



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 Gambaran Umum Objek Penelitian

Dalam penelitian ini mempunyai tujuan untuk memetakan atau memvisualisasikan data gempa bumi yang terjadi pada wilayah di Indonesia, oleh karena itu objek penelitian yang dibutuhkan dalam membangun penelitian ini adalah berupa data gempa bumi di Indonesia yang diambil selama 3 tahun terakhir, yakni selama tahun 2014 hingga 2016. Menurut hasil wawancara dan konsultasi dengan pihak BMKG, diketahui bahwa untuk mengetahui visualisasi berupa pemetaan gempa bumi di Indonesia cukup dengan menggunakan data gempa bumi yang terjadi selama 3 tahun terakhir, karena penelitian ini tidak menganalisa pergerakan lempeng atau struktur tanah di bumi. Berikut ini merupakan data-data yang digunakan dalam pembuatan visualisasi gempa bumi dalam penelitian ini.

3.1.1 Data Gempa Bumi Seluruh Daerah Di Indonesia



Gambar 3.1 Data Online BMKG

Sumber : <http://dataonline.bmkg.go.id>

Penulis melakukan pengambilan data gempa bumi yang terjadi pada seluruh daerah di Indonesia dari pusat *data online* milik BMKG Indonesia. Tujuan penulis dalam mengambil data gempa bumi seluruh daerah di

Indonesia adalah untuk memvisualisasikan seluruh titik gempa bumi yang terjadi di seluruh 34 provinsi di Indonesia secara menyeluruh. Variabel yang dilihat adalah lokasi terjadi gempa, kedalaman gempa, dan kekuatan gempa.

3.1.2 Data Gempa yang Dirasakan

Data ini penulis dapatkan secara *online* dari *website* milik InaTEWS (*Indonesia Tsunami Early Warning System*) yang merupakan bagian dari BMKG.



Gambar 3.2 InaTEWS Logo

Sumber : <http://inatews.bmkg.go.id/new/>

Tujuan penulis mengambil data ini adalah untuk mengetahui skala intensitas gempa yang dapat dirasakan oleh manusia, dimana skala ini dibagi menjadi beberapa kategori nantinya.

3.1.3 Data Gempa Bumi pada Provinsi Nanggro Aceh Darussalam

Penulis juga mengambil data gempa bumi yang terjadi pada provinsi Nanggro Aceh Darussalam secara langsung di kantor pusat BMKG Indonesia yang terletak pada jalan Angkasa I No. 2, Kemayoran, RW.10, Gn. Sahari Sel., Kemayoran, Kota Jakarta Pusat, DKI Jakarta.



Gambar 3.3 Logo BMKG

Sumber : <http://www.bmkg.go.id/profil/?p=logo-bmkg>

Hal ini dilakukan penulis dengan tujuan untuk memvisualisasikan data gempa bumi yang lebih detail pada provinsi yang sering mengalami kerusakan saat terjadi gempa pada wilayah di Indonesia, yakni provinsi Nanggroe Aceh Darussalam. Provinsi Nanggroe Aceh Darussalam terpilih menjadi salah satu daerah yang paling sering mengalami gempa bumi dan kerusakan, hal ini diketahui setelah dilakukannya wawancara dengan Pak Urip Setiyono yang berkerja di kantor pusat BMKG Indonesia bagian mitigasi gempa bumi.

U
M
M
N

3.2 Penelitian Sebelumnya

Dalam menyusun penelitian ini, terdapat beberapa referensi penelitian yang berkaitan dengan visualisasi data dan juga metode *Visual Data Mining*. Penelitian sebelumnya yang pertama kali dibaca adalah sebuah penelitian pada tahun 2005 dengan judul “*Visual Data Mining and Machine Learning*” (Rossi, 2005). Penelitian ini bertujuan untuk mempermudah manusia dalam mengerti dan melihat data yang tidak terstruktur dalam jumlah yang besar. Penelitian ini menjelaskan beberapa tujuan atau tugas dalam penggunaan metode *Visual Data Mining* yakni adalah memperbolehkan pengguna dalam menjelajahi data berjumlah besar dan mencari informasi yang menarik.

Penelitian kedua yang dibaca adalah sebuah penelitian yang berasal dari Jerman dengan judul “*Visual Data Mining Techniques*” (Ward, 2002). Pada penelitian ini Ward mempunyai tujuan untuk mengklasifikasi informasi yang digunakan untuk visualisasi data berdasarkan tipe datanya. Pada penelitian ini Ward mengungkapkan beberapa tipe data yang dapat digunakan untuk visualisasi data adalah tipe data seperti *one-dimensional data*, *two-dimensional data*, *multidimensional data*, *text*, *hypertext*, *hierarchies and graphs*, dan *algorithms and software*.

Penelitian ketiga yang dibaca adalah sebuah penelitian yang terbit pada tahun 2002 dengan judul “*Information Visualization and Visual Data Mining*” (Keim, 2002). Dalam penelitian ini Keim mengambil beberapa contoh data dengan jumlah yang besar dari beberapa kurun tahun lalu, data-data yang telah didapat tersebut selanjutnya diklasifikasi dan divisualisasikan berdasarkan teknik yang

tepat dalam memvisualisasikannya. Penelitian ini juga menjelaskan beberapa teknik yang dapat digunakan dalam memvisualisasikan data berdasarkan jenis datanya seperti *standard 2D/3D displays*, *geometrically transformed*, *icon-based displays*, *dense pixel displays*, dan *stacked displays*.

Penelitian keempat yang dibaca adalah penelitian yang berasal dari *Universitas of Denver* pada tahun 2015 yang berjudul “*Communicating Research via Data Visualization*” (Ryan, 2015). Penelitian ini mengungkapkan banyak informasi yang dapat disampaikan kepada para pendengar ataupun para pengguna dari visualisasi, contohnya dengan menggunakan *line charts*, *tables bar*, dan *map graphics*. Dengan menggunakan model yang sesuai visualisasi, informasi dapat berkomunikasi dengan sendirinya dan memberikan manfaat bagi para penggunanya.

Penelitian terakhir yang telah dibaca adalah penelitian yang terbit pada tahun 2013 yang berjudul “*Improving health and healthcare with interactive visualization methods*” (Shneiderman, 2013). Penelitian ini mengungkapkan bahwa visualisasi dapat meningkatkan kesehatan dan pengobatan manusia dengan cara menggambarkan tingkat kesehatan berdasarkan tingkah laku atau habitat para pasien. Visualisasi dari tingkah laku atau habitat manusia ini dapat menyimpulkan *trend* konsumsi dan pola hidup yang sedang digemari oleh manusia pada saat ini.

3.2.1 Tabel Masalah dan Solusi Penelitian

Tabel 3.1 Masalah dan Solusi Penelitian

No.	Author	Problem	Summary
1	Fabrice Rossi (2005)	Cara mempermudah manusia dalam membaca dan mengerti data yang tidak terstruktur dalam jumlah yang besar	Penggunaan metode <i>Visual Data Mining</i> dalam pembangunan <i>machine learning</i> untuk memvisualisasikan data.
2	Matthew Ward (2002)	Cara mengklasifikasi informasi yang digunakan untuk visualisasi data berdasarkan tipe datanya	Dengan menggunakan metode <i>Visual Data Mining</i> data dapat dibedakan berdasarkan tipe datanya.
3	Daniel A. Keim (2002)	Cara mengklasifikasikan dan memvisualisasikan data yang berasal dari masa lampau	Metode visualisasi data mining dibuat dengan teknik yang tepat dalam memvisualisasikan data tersebut.
4	J. Ryan (2015)	Cara informasi dapat berkomunikasi dengan pada <i>audience</i> melalui visualisasi data	Penggunaan <i>models</i> yang tepat seperti <i>line chart</i> , <i>table bar</i> , dan <i>map graphics</i> dapat membantu komunikasi informasi dengan para <i>audience</i>
5	Ben Shneiderman (2013)	Mengetahui bagaimana cara untuk dapat meningkatkan kesehatan dan pengobatan manusia	Peningkatan kesehatan dan pengobatan manusia dapat dilihat dari simpulan dari visualisasi yang menggambarkan <i>trend</i> konsumsi dan pola hidup dari manusia saat ini

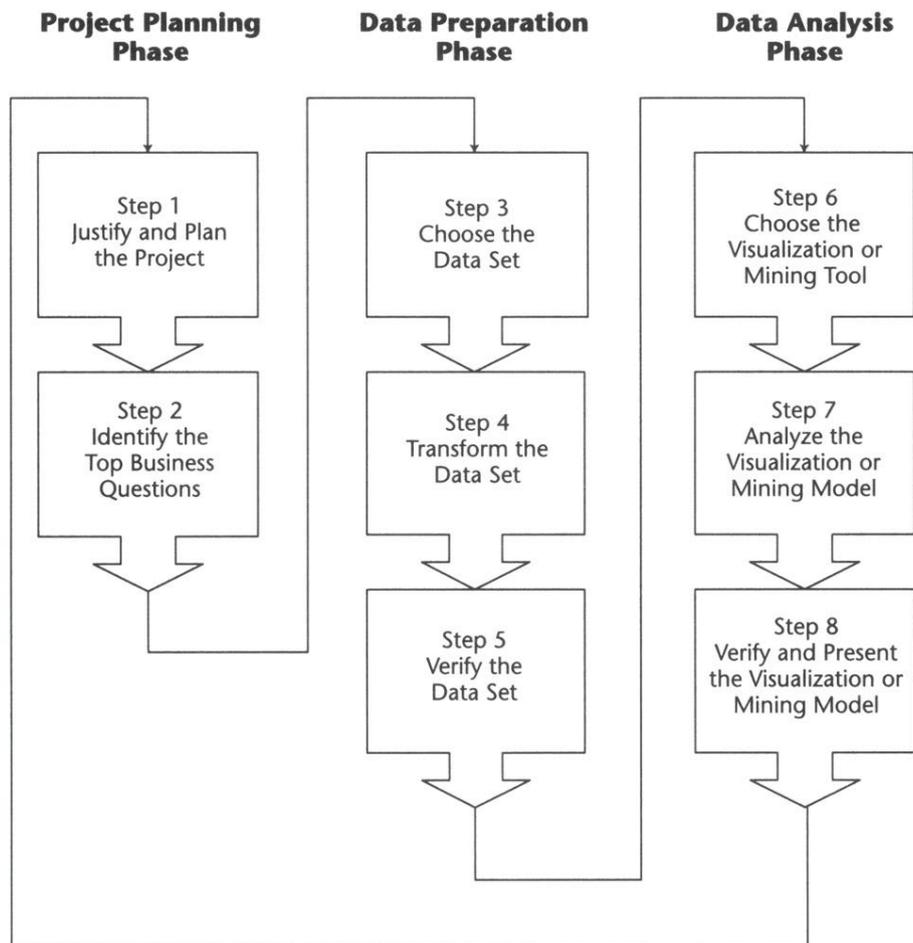
3.3 Metode Penelitian

3.3.1 *Visual Data Mining*

Dalam penelitian ini metode yang digunakan penulis untuk memvisualisasikan data adalah metode *Visual Data Mining* atau yang sering disebut VDM. Penulis menggunakan metode ini karena metode *Visual Data Mining* merupakan sebuah metode yang menyediakan siklus dalam membangun sebuah visualisasi data. Berdasarkan dengan data gempa bumi yang dimiliki oleh penulis maka dalam penelitian ini penulis menggunakan metode *Visual Data Mining* yang dianggap cocok dengan tipe data dan visualisasi data yang akan dibangun nantinya.

Metode *Visual Data Mining* adalah sebuah metode yang terdiri dari beberapa fase dan didalam masing-masing fase tersebut terdapat beberapa langkah yang menyusun satu keutuhan sebuah siklus. Siklus ini dapat dimulai dari persiapan data, pengolahan data, penyimpanan data, dan pemvisualisasian data.

U
M
M
N



Gambar 3.4 *Visual Data Mining Phase and Step*

Sumber : *Visual Data Mining-Techniques and Tools for Data Visualization and Mining* (Soukup, 2002).

Fase dalam metode visual data mining ini dibagi menjadi 3 fase dan terdiri dari 8 langkah, yakni seperti berikut :

A. Project Planning Phase

Pada fase inilah penulis membuat perencanaan proyek dan persiapan kebutuhan dalam menjalankan proyek. Dalam fase ini terdapat 2 langkah yang harus dilakukan, yakni :

1. *Justify and Plan the Project*

Pada langkah ini dilakukan perencanaan mengenai visualisasi apa yang akan dibuat, bagaimana cara membuat visualisasi tersebut, dan apa tujuannya dibuat visualisasi tersebut. Selanjutnya tentukan sumber daya apa yang dibutuhkan untuk membuat penelitian ini, dimana pada penelitian ini membutuhkan sumber daya seperti data gempa bumi dan *tools* yang akan digunakan dalam memvisualisasikan data, seperti *Microsoft Power BI*.

2. *Identify the Top Business*

Pada langkah ini dimaksudkan untuk menentukan kebutuhan pengguna utama visualisasi. Pengguna utama visualisasi yang dibuat oleh penulis adalah Badan Meteorologi, Klimatologi, dan Geofisika. Wawancara juga telah dilakukan secara langsung bersama Pak Urip Setiyono yang berkerja di kantor BMKG Indonesia, hasil wawancara yang dilakukan adalah memperoleh data mengenai wilayah mana di Indonesia yang sering mengalami kerusakan gempa bumi, selain itu penulis juga memperoleh informasi teknis mengenai gempa bumi dan juga pembagian kategori gempa berdasarkan intensitas.

B. Data Preparation Phase

Fase ini digunakan untuk mempersiapkan data-data yang diperlukan, selain itu penggunaan *tools* dan tata cara pengolahan data juga terdapat pada fase ini. Pada fase ini terdapat 3 langkah yakni :

1. *Choose the Data Set*

Pengumpulan data diambil melalui website resmi *data online* milik BMKG Indonesia. Selain itu penulis juga mendapatkan data melalui permintaan data secara langsung di kantor pusat BMKG Indonesia,

2. *Transform the Data set*

Selanjutnya data dalam format *Microsoft Excel* tersebut akan dibersihkan dengan proses *data cleaning* dan Proses ETL, *Tool* yang digunakan untuk proses data cleaning adalah *tools* yang bernama *Open Refine*. *Open Refine* adalah salah satu software *data cleaning* yang dibuat oleh *Google* dan dijalankan pada sistem operasi *windows*. *Google* merancang *software* ini untuk digunakan dalam mengakses data dari jumlah kecil hingga besar, membersihkan dan merubah data, serta melakukan rekonsiliasi dan perbandingan data.

3. *Verify the Data set*

Setelah data dibersihkan dan telah ditransformasi, maka penulis akan mengecek kembali apakah data tersebut sudah benar-benar bersih dari bias dan *error*. *Tool* yang digunakan untuk proses verifikasi ini menggunakan *Pentaho*. *Pentaho* dapat melakukan banyak proses integrasi data secara menyeluruh termasuk melakukan proses ETL dan verifikasi data. Kelebihan *pentaho* dibanding dengan *software* serupa lainnya adalah *GUI* yang ditampilkan oleh *pentaho* yang lebih *user friendly*, selain itu *pentaho* juga dapat terintegrasi dengan *platform business intelligence* miliknya sendiri ataupun *platform business intelligence* lainnya, sehingga lebih mudah untuk digunakan pada lingkup *cross platform*.

C. *Data Analysis Phase*

Penulis memulai fase ini dengan pemilihan *tools* atau alat yang akan digunakan untuk memvisualisasikan data. Pada fase ini data juga mulai divisualisasikan, dianalisis, dan dipresentasikan hasilnya. Di dalam fase ini juga terdapat 3 langkah yang harus dilakukan, yakni :

1. *Choose the Visualization or Mining Tools*

Pada langkah ini, penulis akan membandingkan *tools* yang akan digunakan untuk memvisualisasikan data.

Tool yang digunakan untuk visualisasi data dalam penelitian ini adalah *Microsoft Power BI*. *Microsoft Power BI* merupakan produk dari perusahaan teknologi *Microsoft* dimana produk-produk *Microsoft* seperti *Microsoft office*, *Microsoft excel* dan *access* dapat diakses dengan mudah oleh *Microsoft Power BI*, selain itu juga terdapat banyak kustomisasi visual yang dihadirkan oleh *Microsoft Power BI*.

2. *Analyze the Visualization or Data Mining Model*

Pada langkah ini penulis akan mulai melakukan analisa dari visualisasi data yang telah dibentuk. Bentuk visualisasi data yang dipilih penulis dapat berupa *map graphs* atau *bar graphs* maupun *line graphs*. Penulis akan melakukan analisa terhadap informasi yang ditampilkan pada masing-masing *graphs* yang telah dibentuk..

3. *Verify and Present Analysis*

Pada tahap terakhir ini penulis akan mendapatkan kesimpulan informasi yang telah digali pada tahap analisa sebelumnya. Dimana informasi ini selanjutnya diverifikasi kembali kebenarannya dan juga dipresentasikan kepada pengguna utama yang membutuhkannya, dalam hal ini adalah BMKG.

3.4 Variabel Penelitian

3.4.1 Variabel Independen

Variabel Independen adalah sebuah tipe variabel yang menjelaskan, mempengaruhi variabel lainnya tetapi tidak terpengaruhi oleh variabel lain. Di dalam penelitian ini variabel independen yang digunakan adalah data gempa diseluruh daerah Indonesia dan Nanggroe Aceh Darussalam (Y1) data gempa dirasakan diseluruh daerah Indonesia dan Nanggroe Aceh Darussalam (Y2).

3.4.2 Variabel Dependen

Variabel Dependen adalah tipe variabel yang dijelaskan atau terpengaruhi oleh variabel independen atau lainnya. Dalam penelitian ini variabel dependen yang digunakan adalah, kekuatan gempa (X_1), kedalaman gempa (X_2), dan Intensitas (X_3).

UMMN

3.5 Perbandingan *Tools*

Berikut ini merupakan hasil perbandingan dari *tools* yang nantinya akan menjadi acuan pada penelitian ini dalam memilih *tools* digunakan.

Tabel 3.2 Perbandingan *Tools Cleaning Data*

ASPECT	OPEN REFINE (v.2.5)	WRANGLER (v.1.1)
<i>Integration And Sharing</i>	Dapat terhubung dengan produk-produk dari <i>google</i> seperti <i>google spreadsheet</i> , <i>google fusion tabel</i> .	Tidak dapat terintegrasi dengan <i>software</i> lainnya karena merupakan <i>software</i> yang berdiri sendiri
<i>Supported import</i>	<i>XLS, TSV, CSV, Text File, XML, RDF, JSON, Google Spreadsheet, Google Fusion tabel.</i>	<i>XLS, CSV, JSON, Text File, TSV, LOG, TDE.</i>
<i>Supported Export</i>	<i>XLS, TSV, CSV, HTML Tabel, Templating exporter.</i>	<i>CSV, JSON, TDE.</i>
<i>Costs for full version</i>	<i>Free (Gratis)</i>	<i>Trifacta</i> tidak menyebutkan harga pasti terhadap produknya, tergantung dengan permintaan konsumennya.
CONCLUSION	Dengan mempertimbangkan integrasi, pendukung <i>import</i> dan <i>export</i> , serta dari segi harga, <i>Open Refine</i> lebih diunggulkan daripada pesaingnya, sehingga dalam penelitian ini akan digunakan <i>Open Refine</i> .	

Sumber :

<https://r-dir.com/blog/2012/11/data-wrangler-vs-google-refine.html>

https://groups.google.com/forum/#!topic/openrefine/AWG1hIMcn_o

Tabel 3.3 Perbandingan Tools ETL

ASPECT	PENTAHO (v.3.2)	TALEND (v.4.0)
<i>Integration And Sharing</i>	Dapat terintegrasi dengan produk <i>Pentaho</i> lainnya, selain itu juga dapat masuk ke sistem <i>Hadoop</i> dan beberapa <i>Open Source</i> .	Dapat terintegrasi dengan beberapa <i>software</i> dengan <i>java-based programming</i> .
<i>User Environment</i>	<i>Pentaho</i> lebih mudah dan cepat untuk dibuka dibanding dengan pesaingnya yang lain	<i>Software</i> ini membutuhkan beberapa langkah terlebih dahulu yang diperlukan untuk menggunakannya.
<i>Reliability, maturity, support</i>	Informasi yang tersedia pada komunitas dan forum <i>Pentaho</i> lebih lengkap.	<i>Talend</i> merupakan sebuah <i>software</i> baru, oleh karena itu informasi komunitas dan forumnya terbatas
<i>Components and technology</i>	Dibangun dengan menggunakan <i>java swing application</i> , dijalankan dengan menggunakan <i>javascript</i>	Dibangun dengan basis <i>eclipse java tool</i> , prosedur di jalankan dengan mengkompilasi kode <i>java</i> .
CONCLUSION	Kedua <i>software</i> ETL ini sama-sama memiliki fitur yang unggul dalam melakukan proses ETL. Dalam hal ini, <i>Pentaho</i> lebih unggul dalam pemberian informasi baik oleh forum ataupun komunitasnya, oleh karena itu dalam penelitian ini digunakan <i>Pentaho</i> .	

Sumber :

http://www.robertomarchetto.com/talend_studio_vs_kettle_pentao_pdi_comparis_on

<https://churriwifi.com/2010/06/01/comparing-talend-open-studio-and-pentaho-data-integration-kettle/>

Tabel 3.4 Perbandingan Tools Visualisasi

ASPECT	POWER BI (v.2.37)	TABLEAU (v.9.35)
<i>Data Sources And Modelling</i>	<i>Microsoft Power BI menyediakan machine learning yang dapat digunakan untuk menganalisa data</i>	Tidak memiliki <i>machine learning</i> dan sulitnya untuk melakukan <i>web query</i> dalam pembuatan <i>dashboard</i>
<i>Custom Visuals</i>	<i>Power BI dapat mengambil data template dari komunitasnya ataupun langsung dari google analytics</i>	Visualisasi yang terbatas pada <i>Tableau</i> , serta terbatasnya fitur pengambilan visualisasi dari <i>google analytics</i>
<i>Dax Expressions</i>	<i>Power BI dapat menggunakan formula dax expressions, dimana forumula ini dapat menghitung data, kolom, dan agregasi</i>	<i>Tableau</i> tidak menyediakan formula <i>dax expression</i> seperti pada <i>Power BI</i> .
<i>Cost</i>	Lisensi <i>Power BI</i> dapat dibeli oleh 1 <i>user</i> dengan harga \$12 / bulan atau \$144 / tahun	Lisensi <i>tableau dekstop</i> saat ini berkisar diantara \$1000-2000 / tahun
<i>Depth of discovery</i>	<i>Power BI</i> cocok untuk menyelidiki kumpulan data bisnis dengan menggunakan <i>ms.excel</i>	Fleksibel untuk visualisasi, namun tidak bisa digunakan untuk memanipulasi data.
CONCLUSION	<i>Microsoft Power BI</i> menyediakan banyak fitur visualisasi dengan harga yang lebih murah dibanding <i>tableau</i> dan dapat menggunakan <i>dax expressions</i> .	

Sumber :

<https://www.betterbuys.com/bi/tableau-vs-power-bi/>

<https://comparisons.financesonline.com/microsoft-power-bi-vs-tableau-software>

<https://selecthub.com/business-intelligence/tableau-vs-qlikview-vs-microsoft-power-bi/>