**VISUALISASI DATA INTERAKTIF DATA *“OPEN DATA”* EKONOMI DAN KEUANGAN DAERAH PEMERINTAH PROVINSI DKI JAKARTA**

**SKRIPSI**

Disusun untuk Memenuhi Ujian Akhir Sarjana

**NADIAR AHMAD SYARIPUL  
10111121**



**PROGRAM STUDI TEKNIK INFORMATIKA**

**FAKULTAS TEKNIK DAN ILMU KOMPUTER**

**UNIVERSITAS KOMPUTER INDONESIA**

**2015**

# ABSTRAK

# ABSTRACK

# DAFTAR ISI

[ABSTRAK i](#_Toc434339005)

[ABSTRACK ii](#_Toc434339006)

[DAFTAR ISI iii](#_Toc434339007)

[DAFTAR GAMBAR v](#_Toc434339008)

[DAFTAR TABEL vi](#_Toc434339009)

[DAFTAR LAMPIRAN vii](#_Toc434339010)

[BAB I PENDAHULUAN 1](#_Toc434339011)

[I.1 Latar Belakang Masalah 1](#_Toc434339012)

[I.2 Perumusan Masalah 2](#_Toc434339013)

[I.3 Maksud dan Tujuan 2](#_Toc434339014)

[I.4 Batasan Masalah 2](#_Toc434339015)

[I.5 Metodologi Penelitian 2](#_Toc434339016)

[I.5.1 Metode Pengumpulan Data 3](#_Toc434339017)

[I.5.2 Metode Pengembangan Perangkat Lunak 3](#_Toc434339018)

[I.6 Sistematika Penulisan 4](#_Toc434339019)

[BAB II LANDASAN TEORI 6](#_Toc434339020)

[II.1 Konsep Data 6](#_Toc434339021)

[II.1.1 Data 6](#_Toc434339022)

[II.1.2 Basis Data 6](#_Toc434339023)

[II.2 Visualisasi Data 7](#_Toc434339024)

[II.2.1 Tipe-Tipe Visualisasi 7](#_Toc434339025)

[II.2.2 Proses Visualisasi Data 11](#_Toc434339026)

[II.3 Descriptive Statistics 12](#_Toc434339027)

[II.3.1 Distribusi Frekuensi dan *Percentile Rank* 13](#_Toc434339028)

[II.3.2 Ukuran Gejala Pusat dan Variasi 13](#_Toc434339029)

[II.3.3 *z* Skor dan Distribusi Normal 15](#_Toc434339030)

[II.4 Inferential Statistics 15](#_Toc434339031)

[II.4.1 Probabilitas 16](#_Toc434339032)

[II.4.2 Sampling Distribution 16](#_Toc434339033)

[II.4.3 Hypothesis Testing 16](#_Toc434339034)

[II.4.4 Parameter Estimation 16](#_Toc434339035)

[II.5 Algoritma 16](#_Toc434339036)

[BAB III ANALISIS DAN PERANCANGAN DATA 17](#_Toc434339037)

[III.1. Analisis Sistem 17](#_Toc434339038)

[III.1.1 Analisis Masalah 17](#_Toc434339039)

[III.1.2 Analisis Sumber Data 17](#_Toc434339040)

[III.1.3 Analisis Segmentasi Pengguna Informasi 17](#_Toc434339041)

[III.1.4 Analisis Kebutuhan Informasi Berdasarkan Segmentasi Pengguna 17](#_Toc434339042)

[III.1.5 Pemetaan Visualisasi Data 17](#_Toc434339043)

[III.1.6 Analisis Algoritma 17](#_Toc434339044)

[III.1.7 Proses Visualisasi Data 17](#_Toc434339045)

[DAFTAR PUSTAKA 1](#_Toc434339046)

# DAFTAR GAMBAR

# DAFTAR TABEL

# DAFTAR LAMPIRAN

# PENDAHULUAN

## Latar Belakang Masalah

Pada era digital seperti sekarang, menemukan data tidaklah begitu sulit, data (dataset) banyak ditemukan di internet. Dataset tersebut sangatlah besar, kompleks, dan sulit dianalisis dengan perangkat lunak *data processing* seperti *spreadsheet*, dataset seperti ini dikenal dengan istilah *“big data”* [1]. Pemerintah Provinsi DKI Jakarta menyediakan dataset yang bebas didistribusikan, disimpan, dan dikelola. Dataset seperti ini dikenal dengan istilah *“open data”* [2]. Saat penelitian ini dilakukan, dataset yang terdapat pada repository *“open data”* Pemerintah Provinsi DKI Jakarta sebanyak 556 dataset. Selain sebagai transparansi, dataset itu juga sebagai upaya Pemerintah Provinsi DKI Jakarta menyediakan satu basis data pembangunan yang akurat, terbuka, terpusat dan terintegrasi [3].

Berdasarkan hasil kuesioner terbuka yang dilakukan pada tanggal 25 Agustus 2015 s/d 27 Agustus 2015 dengan segmentasi responden pelaku UKM atau *entrepreneur*, mayoritas responden setuju dengan transparansi data (29/39), tetapi menilai bahwa data yang disajikan dalam repositori masih sulit dibaca dan dimengerti karena format data yang tidak familiar dan terpisah-pisah (27/39). Sangat sulit bagi responden untuk mengambil gambaran umum dari data yang tersedia.

Cara efektif menyajikan data yang detail menjadi informasi yang mudah diterima adalah dengan cara abstraksi menjadi informasi visual. Visualisasi melalui perumpamaan visual dari dulu hingga saat ini telah digunakan sebagai teknik penyimpanan pesan. Dahulu perumpamaan visual berbentuk simbol-simbol yang merepresentasikan suatu makna. Visualisasi mengubah data menjadi informasi yang bisa dimengerti secara universal [4].

Berdasarkan masalah yang dipaparkan, dapat disimpulkan bahwa publik memerlukan visual analisis dari dataset tersebut. Visualisasi interaktif dapat menjadi solusi mayoritas publik dalam memahami gambaran umum dari dataset yang ada.

## Perumusan Masalah

Berdasarkan pemaparan latar belakang masalah, maka perumusan permasalahan yang terdapat pada penelitian ini adalah memvisualisasikan dataset ekonomi dan keuangan daerah Pemerintah Provinsi DKI Jakarta secara interaktif.

## Maksud dan Tujuan

Maksud dari penelitian ini adalah memvisualisasikan dataset ekonomi dan keuangan Pemerintah Provinsi DKI Jakarta. Adapun tujuan yang penelitian ini adalah sebagai berikut:

1. Membantu memudahkan publik melihat gambaran umum dataset ekonomi dan keuangan daerah dengan cara membuat visualisasi interaktif.

## Batasan Masalah

Adapun batasan masalah dari penelitian ini adalah:

1. Dataset utama yang digunakan adalah dataset keuangan daerah, perdagangan, dan perekonomian yang didapatkan dari repositori *open data* Pemerintah Provinsi DKI Jakarta.
2. Pendekatan analisis perangkat lunak menggunakan analisis terseruktur.
3. Segmentasi pengguna produk dari penelitian ini adalah pelaku UKM dan *entrepreneur*.
4. Sistem hanya menampilkan informasi visualisasi interaktif, tidak menerima data personal pengguna.

## Metodologi Penelitian

Metodologi penelitian yang digunakan dalam penelitian ini menggunakan metode penelitian terapan, metode ini bertujuan untuk menyelesaikan permasalahan berdasarkan teori-teori yang sudah ada [5]. Adapun metodologi yang digunakan dalam penelitian ini menggunakan dua metode, yaitu metode pengumpulan data dan metode pembangunan perangkat lunak.

## Metode Pengumpulan Data

Adapun teknik pengumpulan data yang akan digunakan terdiri dari tiga cara pengumpulan data, yaitu:

1. Studi Literatur

Studi literatur utama dari penelitian ini bersumber dari buku “*Data Analysis with Open Source Tools*” dari K. Janert, “*Doing Data Science*” dari Rachel Schutt & Cathy O’Neil, *Massive Open Online Course* (MOOC) Udacity: Data Analyst Naonodgree, serta beberapa MOOC dan jurnal lainnya.

1. Kuesioner

Kuesioner dilakukan dengan memberikan beberapa pertanyaan kepada calon responden dari grup facebook UKM atau *Entrepreneur* melalui fasilitas google form.

1. Pencarian Dokumen

Dokumen yang digunakan adalah dokumen yang bersumber dari repositori *open data* Pemerintah Provinsi DKI Jakarta.

## Metode Pengembangan Perangkat Lunak

Dalam pembangunan perangkat lunak ini, menggunakan model *waterfall* sebagai tahapan pengembangan perangkat lunak [6]. Adapun proses tersebut antara lain:

1. *Requirenent Definition*

Tahap *requirement definition* adalah tahap dimana kebutuhan sistem, konstrain, dan tujuan pembangunan sistem didefinisikan dengan detail.

1. *System and Software Design*

Tahap *System and Software Design* merupakan tahap proses perancangan sistem yang meliputi identifikasi kebutuhan fungsional beserta relasinya. Tahap ini dapat dikerjakan setelah kebutuhan perangkat keras dan lunak telah terpenuhi.

1. *Implementation and Unit Testing*

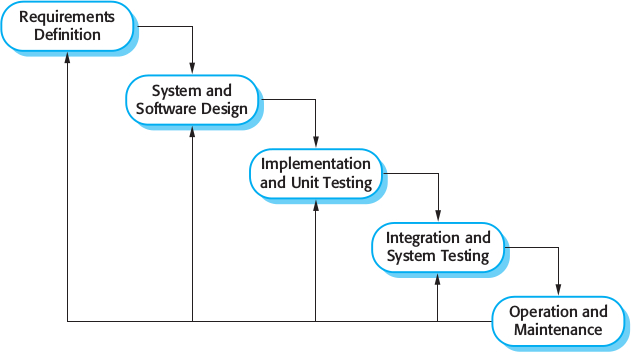
Tahap *Implementation and Unit testing* merupakan tahap dimana perancangan desain yang telah dibuat, diterjemahkan kedalam kode-kode unit program menggunakan bahasa pemrograman. Setelah itu sistem yang sudah jadi diuji apakah setiap unit mencapai spesifikasinya.

1. *Integration and System Testing*

Tahap *integration and system testing* merupakan tahap dimana setiap unit program diintegrasikan dan diperiksa secara keseluruhan untuk memastikan apakah sudah mencapai kebutuhan yang telah ditentukan.

1. *Operation and Maintentance*

Tahap *operation and maintenance* merupakan tahap pemeliharaan berkelanjutan. *Maintenance* meliputi melakukan perbaikan terhadap *bugs-bugs* yang ditemukan dikemudian hari, melakukan pengembangan perangkat lunak, dan menambahkan fungionalitas baru ketika kebutuhan baru diperlukan.



Gambar I‑1 Model *Waterfall* [6]

## Sistematika Penulisan

Sistematika penulisan yang akan digunakan untuk laporan akhir penelitian disusun untuk memberikan gambaran umum tentang penelitian yang dikerjakan. Sistematika penulisan dalam tugas akhir ini adalah sebagai berikut:

**BAB 1 PENDAHULUAN**

Bab 1 secara singkat membahas tentang latar belakang permasalahan, merumuskan inti permasalahan, mencari solusi atas masalah yang muncul, melakukan identifikasi masalah tersebut, menentukan maksud dan tujuan, kegunaan penelitian, pembatasan masalah, metode penelitian, dan sistematika penulisan.

**BAB 2 LANDASAN TEORI**

Bab 2 menguraikan bahan-bahan kajian, konsep dasar, dan teori dari para ahli yang berkaitan dengan penelitian. Meninjau permasalahan dan hal-hal yang berguna dari penelitian-penelitian dan sintesis serupa yang pernah dikerjakan sebelumnya dan menggunakannya sebagai acuan pemecahan masalah pada penelitian ini.

**BAB 3 ANALISIS DAN PERANCANGAN SISTEM**

Bab 3 menguraikan hasil analisis dari objek penelitian untuk mengetahui hal atau masalah apa yang timbul dan mencoba memecahkan masalah tersebut dengan mengaplikasikan perangkat-perangkat dan pemodelan yang digunakan.

**BAB 4 IMPLEMENTASI DAN PENGUJIAN SISTEM**

Bab 4 menguraikan tentang perancangan solusi beserta implementasinya dari masalah-masalah yang telah dianalisis. Pada bagian ini juga akan ditentukan bagaimana sistem dirancang, dibangun, diuji dan disesuaikan dengan hasil penelitian.

**BAB 5 KESIMPULAN DAN SARAN**

Bab 5 menguraikan tentang kesimpulan dari hasil penelitian beserta saran untuk pengembangan selanjutnya.

# LANDASAN TEORI

## Konsep Data

### Data

Menurut kamus internasional Cambridge data adalah kumpulan dari fakta-fakta yang berupa angka yang digunakan untuk membantu pengambilan keputusan. Menurut Joseph M. Firestone dalam buku *“Enterprise Information Portals and Knowladge Management”* data adalah kumpulan dari datum, yang mana datum adalah nilai observasi yang bisa diukur atau dihitung. Kesimpulannya, data merupakan fakta-fakta dari fenomena yang sudah ada.

### Basis Data

Basis data adalah kumpulan data yang merepresentasikan dunia nyata di disain dan dibangun dengan tujuan tertentu [7].

Perangkat lunak yang membantu pengguna dalam mengelola dan membuat basis data disebut dengan istilah DBMS (*Data Base Management System*) [8]. Interaksi antara pengguna dan DBMS ditentukan melalui bahasa atau sintak khusus sesuai dengan jenis DBMS yang digunakan. Pada umumnya bahasa DBMS dikelompokan ke dalam dua bentuk, yaitu DDL dan DML [8].

1. DDL (*Data Definition Language*)

DDL adalah bahasa untuk membuat tabel, membuat indeks, mengubah tabel, dan menentukan struktur penyimpanan tabel. Keluaran dari perintah DDL adalah kumpulan tabel yang disimpan dalam berkas khusus yang disebut *data dictionary*.

1. DML (*Data Manipulation Language*)

DML adalah bahasa yang digunakan untuk melakukan manipulasi dan pengambilan data pada suatu basis data. Manipulasi data dapat berupa: penambahan data baru ke basis data; penghapusan data dari basis data; pengubahan data pada basis data.

## Visualisasi Data

Data yang sangat banyak, kompleks, dan tidak teratur sangat sulit untuk dimengerti sekilas pandang oleh manusia. Membaca data akan menjadi masalah ketika data secara inkremental bertambah atau berubah. Tujuan visualisasi data adalah menunjukan mana hal-hal yang penting, melihat pola, dan menunjukan informasi dari berbagai dimensi [9].

Visualisasi data tidak semerta-merta mengubah data menjadi grafik visual, visualisasi data memerlukan perencanaan. Setiap jenis data memiliki teknik visualisasi berdasarkan kebutuhannya [9]. Berdasarkan tingkat kompleksitas data, untuk menghasilkan solusi yang berharga diperlukan dari berbagai disiplin ilmu, yaitu: statistika, *data mining*, desain grafis, dan visualisasi informasi [9].

### Tipe-Tipe Visualisasi

Taksonomi visualisasi dibedakan berdasarkan seberapa dalam abstraksi informasi yang disembunyikan. Terdapat tujuh hal yang harus dipenuhi dalam melakukan abstraksi tingkat tinggi (*high-level abstraction*), tujuh hal tersebut yaitu [10]:

1. *Overview*: Melihat gambaran dari keseluruhan data;
2. *Zoom*: Memperbesar *item* yang terlihat menarik.
3. *Filter*: Melakukan penyaringan terhadap *item* yang dirasa kurang menarik.
4. *Details-on-demand*: Pilih satu *item* dari grup tertentu dan dapat melihat detail kapan saja.
5. *Relate*: Lihat relasi dari setiap *item*.
6. *History*: Dapat mengulang kembali atau kembali ke aksi sebelumnya.
7. *Extract*: Dapat melakukan ekstraksi dari parameter yang diberikan.

Berdasarkan taksonominya, grafik visual dibedakan menjadi [10]: 1D/Linear; 2D/Planar; 3D/Volumetric; Temporal; Multidimensional; Tree/Hierarchical; dan Network.

#### 1D/Linear

Grafik 1-dimensi termasuk di dalamnya adalah tipe data tekstual, kode sumber program, dan huruf alfabet. Setiap *item* yang digambarkan memiliki elemen garis . Contoh dari grafik 1D seperti kode-kode DNA, perbedaan kode sumber, dan lain-lain.

#### 2D/Planar

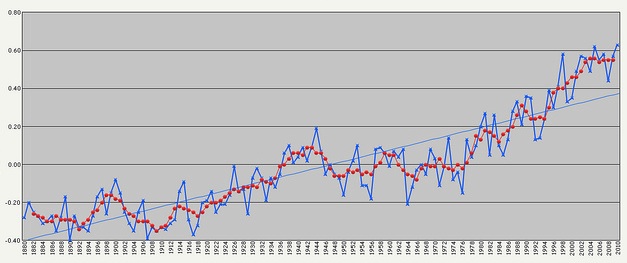
Grafik 2-dimensi termasuk di dalamnya peta geografis, denah rancangan, atau *layout* koran. Setiap *item* pada grafik 2-dimensi memiliki total area dan atribut (warna, ukuran, dll).

#### 3D/Volumetric

Grafik 3-dimensi adalah visual yang menggambarkan objek nyata, seperti tubuh manusia, bentuk bangunan, dll. Setiap *item* pada grafik 3-dimensi memiliki volume.

#### Temporal

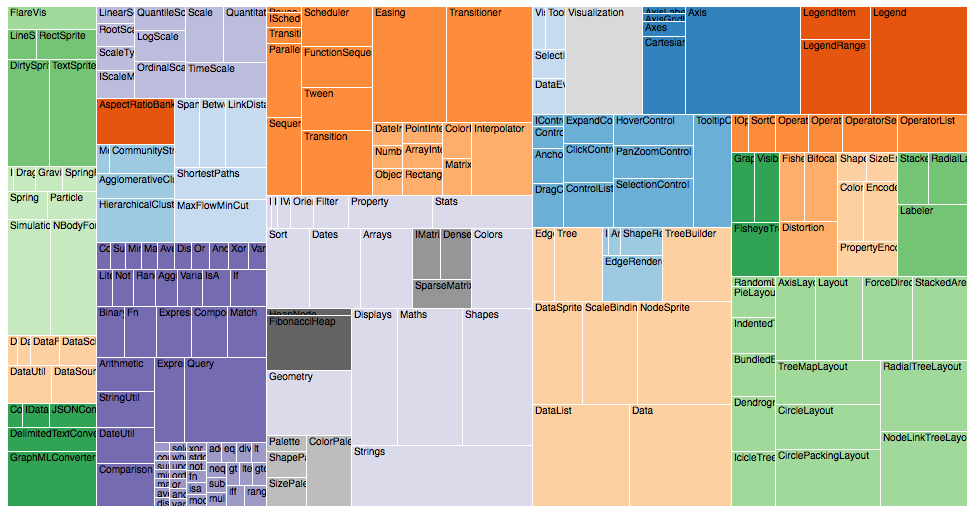
Grafik temporal adalah grafik yang berhubungan dengan waktu (*time lines*). Grafik ini menggambarkan persentasi historikal dari data 1-dimensi. Yang membedakan, grafik temporal memiliki *item* dengan waktu awal dan waktu akhir, atau periode tertentu.



Gambar II‑1 Contoh Grafik Temporal [11]

#### Multi-dimensional

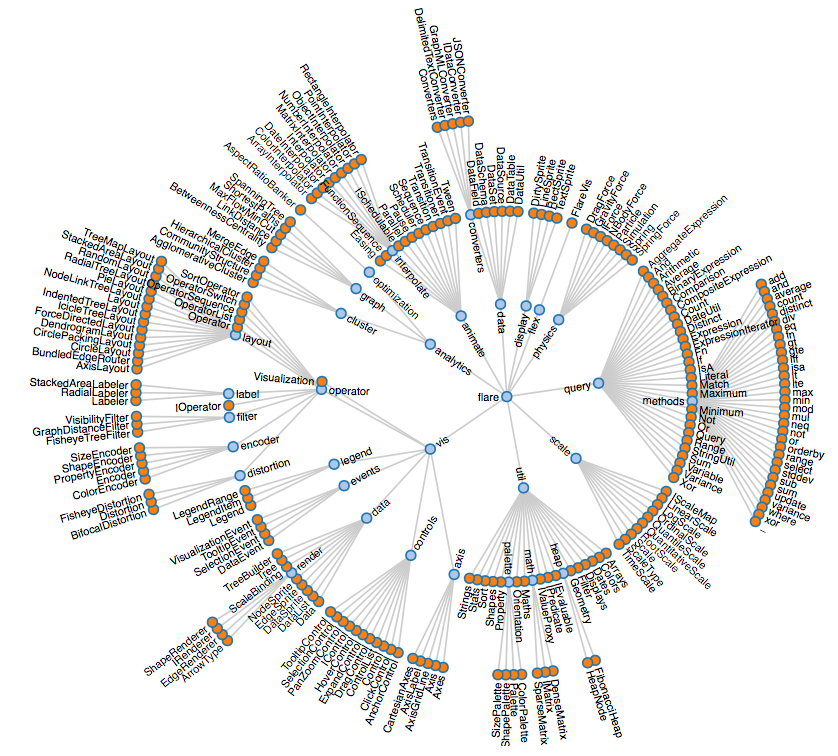
Grafik temporal didalamnya termasuk grafik-grafik yang dihasilkan dari manipulasi data dari disiplin ilmu statistika. Antarmuka representasi milti-dimensional adalah grafik 2-dimensi. Grafik multi-dimensi termasuk didalamnya grafik *pie*, histogram, *tag cloud*, *bubble cloud*, bar, *tree-map*, *scatter plot*, *bubble chart, line chart*, *step chart*, *heat-map*, *parallel sets, spider chart*, *box-plot*, *mosaic display*, *waterfall*, dan tabular.



Gambar II‑2 Tree-map (Multi-dimensi) [11]

#### Tree/Hierarchical

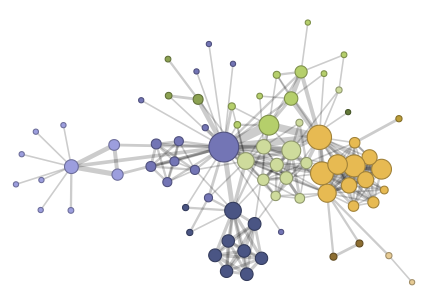
*Tree* adalah grafik herarkikal dari item-item yang memiliki hubungan satu dengan lainnya, atau yang memiliki induk (kecuali *root*). Setiap *item* antara induk dan anak bisa memiliki banyak atribut. Grafik *tree* termasuk didalamnya grafik *tree, dendorogram, radial-tree, hyperbolic-tree, tree-map*, dan *sunburst*.



Gambar II‑3 Contoh Radial-tree [11]

#### Network

*Network* adalah grafik relasi yang tidak bisa digambarkan oleh grafik *tree*, *network* memvisualisasikan relasi-relasi *item* yang banyak (semena-mena). Grafik *network* termasuk di dalamnya grafik matrix, *node-diagram*, *dependency*, *hive-plot*, *alluvial diagram*, dan *subway*.



Gambar II‑4 Contoh Node-diagram [11]

### Proses Visualisasi Data

Visualisasi data sama halnya seperti berkomunikasi, berhasil atau tidaknya komunikasi ditentukan oleh bagaimana cara pembicara menyampaikan informasi yang diberikan kepada lawan komunikasi. Terlalu meluas apa yang dibicarakan juga termasuk hal yang buruk dalam berkomunikasi. Visualisasi yang bagus tentu terfokus, memberikan jawaban yang jelas, dan tidak terlalu detail [9].

Proses memahami sebuah data dimulai dari beberapa pertanyaan. Selanjutnya tidak semerta-merta dijawab begitu saja, tetapi terdapat langkah-langkah dalam menjawab pertanyaan berdasarkan data. Langkah-langkah tersebut yaitu [9]:

1. *Acquire*

Tahap ini adalah tahap dimana data dikumpulkan dari berbagai sumber. Tahap ini bisa menjadi sangat rumit (contoh: mengambil data dari berkas log sistem yang besar). Tahap *acquire* hanya peduli bagaimana data didapatkan, jika produk akhir akan didistribusikan melalui internet, maka data yang ada harus memiliki struktur yang bisa disimpan dalam suatu server.

1. *Parse*

Tahap ini adalah tahap dimana data dekemas ke dalam bentuk atau format data baku dan dikelompokan ke dalam beberapa kategori. Setiap berkas data harus dapat dibaca dan bisa dibedakan dengan data lain.

1. *Filter*

Pada tahap ini beberapa data yang tidak terlalu penting dengan pertanyaan yang akan dijawab dihilangkan. Beberapa data pada berkas, mungkin perlu diterjemahkan ke dalam model matematika atau dilakukan normalisasi terlebih dahulu.

1. *Mine*

Pada tahap ini, data diaplikasikan ke dalam metode disiplin ilmu statistika atau *data mining* sebagai jalan untuk mencari pola atau dijabarkan pada konteks matematis.

1. *Represent*

Pada tahap ini data diubah dan disajikan dalam bentuk model visual dasar, seprti bar, *list*, atau *tree*. Tahap *represent* menunjukan bentuk dasar data yang akan diambil. Tahap ini adalah tahap yang sangat penting dalam membuat visualisasi data. Pemilihan model visualisasi yang tepat dapat memengaruhi bagaimana nantinya kualitas produk yang akan dihasilkan.

1. *Refine*

Pada tahap ini, model visual dibuat lebih jelas lagi dan lebih menarik. Grafik desain lebih banyak berbicara pada tahap ini. Titik-titik yang cukup penting pada visual grafik dibandingkan titik lainnya diberikan pembeda agar dapat mudah dibaca.

1. *Interact*

Pada tahap *interact*, metode untuk memanipulasi data ditambahkan, dengan arti bahwa data bisa ditampilkan sesuai kehendak pengguna. Contoh interaksi antara pengguna dan data seperti *zoom-in*, *zoom-out*, merubah rentang data, melakukan *filtering*, dll.

## Descriptive Statistics

*Descriptive statistics* adalah suatu metode untuk mengorganisasikan dan melihat gambaran umum dari pengukuran sampel dan populasi. Metode *descriptive statistics* seperti namanya biasa dipakai untuk membaca data, seringkali digunakan pada sebuah sampel populasi karena sulitnya untuk mendapatkan seluruh data populasi [12]. *Descriptive statistics* mencakup metode untuk menyusun dan mengelompokan distribusi data dan termasuk mencari satu nilai yang dapat menggambarkan semua nilai dari sampel atau populasi. Menurut Arthur M. Glenberg dan Matthew E. Andrzejewski dalam bukunya yang berjudul *“Learning from Data an Introduction to Statistical Reasoning 3rd”* terdapat tiga teknik penting pada *descriptive statistics*, distribusi frekuensi dan presentil, titik tengah dan variabilitas, dan z skor dan distribusi normal.

### Distribusi Frekuensi dan *Percentile Rank*

Distribusi frekuensi adalah tabulasi jumlah kemunculan dari setiap nilai. Distribusi frekuensi mengelola dan memberikan gambaran umum data, yaitu dengan cara menampilkan karakteristik terbanyak [12]. *Percentile rank* adalah persentasi skor nilai yang didapatkan dari perhitungan distribusi di bawah skor nilai. *Percentile rank* dan persen tidaklah sama. Skor nilai sendiri sering disebut sebagai *percentile* (*P*th), adalah skor nilai *P*% dari perhitungan di bawah distribusi [12].

### Ukuran Gejala Pusat dan Variasi

Ukuran gejala pusat distribusi adalah nilai yang dekat dengan titik tengah dari distribusi. Biasanya digunakan sebagai representasi nilai distribusi keseluruhan [12].

#### Notasi Sigma

∑ adalah karakter kapital dari huruf Yunani yang disebut dengan *sigma*. ∑ digunakan sebagai persamaan jumlah dalam bentuk lebih pendek. Contoh: terdapat sebuah data pelanggaran dari sebuah aktivitas olah raga [7], berikut adalah data tersebut:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Atlit** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| **Jumlah Pelanggaran (X)** | 0 | 3 | 1 | 0 | 2 | 2 | 2 | 5 | 1 |

Dalam notasi sigma bisa ditulis sebagai berikut:

*X* adalah variable yang dihitung, dalam contoh di atas adalah “jumlah pelanggaran”. *i* menunjukan posisi index ke-*i*, *i*=1 di bawah ∑ menunjukan index posisi pertama yang akan dijumlahkan, *n* di atas ∑ menunjukan index posisi terakhir yang akan dihitung. Sehingga expresi sigma di atas bisa ditulis seperti:

Dari data dan persamaan di atas, dapat diketahui bahwa jumlah pelanggaran dapat ditulis dengan notasi = 16.

#### Ukuran Gejala Pusat

Tedapat tiga perhitungan yang sering digunakan dalam mengukur gejala pusat, rata-rata (*mean*), median, dan modus (*mode*). Nilai pada rata-rata. Median, dan modus tidak dapat mewakili seluruh distribusi, tetapi bisa memberikan gambaran umum.

Rata-rata distribusi didefinisikan sebagai nilai penjumlahan setiap nilai yang terdapat pada distribusi (∑) dibagi dengan jumlah nilai yang terdapat pada distribusi (*n*). Formula yang dipakai untuk menghitung rata-rata adalah sebagai berikut:

Rata-rata =

Rata-rata pada sampel dan populasi dibedakan, untuk rata-rata sampel, dinotasikan sebagai *M*, sedangkan pada populasi dinotasikan sebagai *µ* (mu).

|  |  |
| --- | --- |
| *µ* = | *M* = |

Median distribusi adalah nilai yang berada pada titik tengah distribusi. Formula median dapat dibedakan untuk jumlah distribusi genap dan ganjil.

|  |  |
| --- | --- |
| *Median*ganjil = *X( ½ \* (n + 1))* | *Median*genap = *½( X(n/2) + X(n/2 + 1) )* |

Modus distribusi adalah nilai pada distribusi yang paling sering muncul atau yang memiliki frekuensi terbanyak.

#### Variabilitas

Variabilitas adalah pengukuran beda nilai antara satu nilai distribusi dengan nilai distribusi lain. Pengukuran yang lebih menggambarkan variabilitas adalah varian, varian menghitung variabilitas berdasarkan nilai beda dengan *µ* [12].

Varian dari sebuah populasi dapat dihitung dengan cara [12]: (a) Hitung setiap nilai distribusi dengan *µ*; (b) Kuadratkan setiap nilai yang sudah dihitung; (c) Bagi jumlah dari setiap perhitungan dengan populasi, *N*.

Varian dari sebuah sampel untuk menggambarkan populasi, dihitung dengan formula sebagai berikut:

### *z* Skor dan Distribusi Normal

*z* skor atau disebut juga dengan skor standar, sama seperti persentil, *z* skor juga menghitung posisi relatif. Index *z* skor adalah posisi distribusi relatif terhadap distribusi lain. *z* skor sangat berguna pada distribusi normal [12].

*z* skor adalah skor yang sudah dilakukan standarisasi dengan menjadikan 0 sebagai mean dari distribusi dan 1 sebagai standar deviasi [12]. Formula untuk menghitung *z* skor adalah sebagai berikut:

## Inferential Statistics

Tahap pertama memahami data adalah memahami sampel, selanjutnya, sampel yang diambil seringkali digunakan untuk menyimpulkan kondisi populasi sesungguhnya dimana sampel tersebut diambil. *Inferential Statistics* menggunakan sembarang sampel dari populasi yang diuji untuk membuat kesimpulan terhadap populasi [12]. Inferensial statistik diperlukan karena terdapat dua masalah yang saling berhubungan. Pertama, skor dari populasi selalu dibutuhkan. Kedua, populasi biasanya sangat besar, menghitung skor populasi seringkali tidak bisa dilakukan.

Prosedur inferensial statistik dibedakan menjadi dua kategori, estimasi parameter dan *hypothesis-testing*.

* Estimasi parameter melakukan estimasi parameter populasi dari sampel yang diambil secara acak.
* *Hypothesis-testing* membandingkan dua hipotesis tentang populasi. Data dari sampel acak digunakan untuk menentukan hipotesis mana yang cenderung benar.

### Probabilitas

Konsep dari probabilitias adalah menunjukan seberapa banyak keluaran yang mungkin terjadi dari sampling acak populasi. Sampling probabilitas adalah jumlah yang menunjukan sebarapa banyak event yang muncul dari setiap satu observasi populasi (n = 1).

…

### Sampling Distribution

*Sampling distribution* adalah tipe spesial probabilitas distribusi yang sangat penting pada *inferential statistics.* Secara formal, *sampling distribution* adalah probabilitas distribusi yang dihitung dari semua kemungkinan acak sampel dengan ukuran yang sama dari populasi yang sama [12].

…

### Parameter Estimation

Afasfaf

## Algoritma

Supervised Learning

… regression

…

Unsupervised Learning

... Classification

# ANALISIS DAN PERANCANGAN DATA

## Analisis Sistem

Analisis visualisasi data yang dilakukan yaitu dengan cara menerapkan analisis masalah, analisis kebutuhan informasi, analisis kebutuhan fungsional dan non-fungsional, perancangan data, dan analisis proses visualisasi data.

### Analisis Masalah

Berikut merupakan analisis permasalahan yang terdapat pada repositori *open data* Pemerintah Provinsi DKI Jakarta berdasarkan sudut pandang publik:

~~Membantu memudahkan publik melihat gambaran umum dataset ekonomi dan keuangan daerah dengan cara membuat visualisasi interaktif.~~

1. Sulitnya melihat gambaran umum data yang tersedia di repositori *open data* Pemerintah Provinsi DKI Jakarta menurut sudut pandang publik (sampel yang diambil adalah segmentasi pelaku UKM atau *entrepreneur*).

### Analisis Sumber Data

Sumber data yang akan di analisis berasal dari data repositori *open data* Pemerintah Provinsi DKI Jakarta. Format data yang terdapat di repositori *open data* adalah csv, xls/xlsx, dan json.

Tabel III.1 Contoh Tabel "ABBD DKI Jakarta 2014"

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| nomor\_urut | pos\_anggaran | sumber\_pos\_anggaran | penjelasan\_pos\_anggaran | nilai\_anggaran |
| 4.1.1 | Pendapatan Daerah | Pendapatan Asli Daerah | Pajak Daerah | 3.25E+13 |
| 4.1.2 | Pendapatan Daerah | Pendapatan Asli Daerah | Retribusi Daerah | 1.75E+12 |
| 4.1.3 | Pendapatan Daerah | Pendapatan Asli Daerah | Hasil Pengelolaan Kekayaan Daerah Yang Dipisahkan | 4.48E+11 |
| 4.1.4 | Pendapatan Daerah | Pendapatan Asli Daerah | Lain-Lain Pendapatan Asli Daerah | 4.85E+12 |
| 4.2.1 | Pendapatan Daerah | Dana Perimbangan | Dana Bagi Hasil Pajak/Bagi Hasil Bukan Pajak | 1.77E+13 |
| 4.2.2 | Pendapatan Daerah | Dana Perimbangan | Dana Alokasi Umum | 8.60E+10 |
| 4.3.1 | Pendapatan Daerah | Lain-Lain Pendapatan Daerah Yang Sah | Pendapatan Hibah | 5.00E+12 |

### Analisis Segmentasi Pengguna Informasi

Sampel yang digunakan dalam penelitian ini berjumlah **XX** responden, dengan segmentasi berdasarkan jenis pekerjaan. Adapun jenis pekerjaan responden dibedakan menjadi tiga, sebagai *entrepreneur*, sebagai *tech-savy*, dan sebagai analis pasar. Kuesioner dilakukan selama empat hari dari tanggal 30 Oktober 2015 sampai 3 November 2015.

…

### Analisis Kebutuhan Informasi Berdasarkan Segmentasi Pengguna

… (ex: sebaran harga dll.)

### Pemetaan Visualisasi Data

??? jenis visualisasi, tipe2 ??? visualisasi untuk segmen

Data yg dilibatkan untuk informasi dari visual data

### Analisis Algoritma

Kasus yang tidak bias langsung di generate, algoritma yg dipakai yg untuk mengolah data jadi informasi.

### Proses Visualisasi Data

1, Identifikasi stackholder??

Stackholder: 1. Fakta 2. **Tentukan *early adaptor user*.**

# DAFTAR PUSTAKA

[1] C. Snijders, U. Matzat, and U. Reips, “‘Big Data’: Big Gaps of Knowledge in the Field of Internet Science,” *Int. J. Internet Sci.*, vol. 7, no. 1, pp. 1–5, 2013.

[2] “Open Definition - Defining Open in Open Data, Open Content and Open Knowledge.” [Online]. Available: http://opendefinition.org/. [Accessed: 03-Oct-2015].

[3] “Tentang - Data.jakarta.go.id.” [Online]. Available: http://data.jakarta.go.id/about. [Accessed: 03-Oct-2015].

[4] I. J. Asmara, E. Achelia, W. Maulana, R. Wijayanti, and Y. Rianto, “Teknik Visualisasi Grafik Berbasis Web Di Atas Platform,” *Semin. Nas. Apl. Teknol. Inf. 2009 (SNATI 2009)*, vol. 2009, no. Snati, pp. 44–47, 2009.

[5] M. Berndtsson, J. Hansson, B. Olsson, and B. Lundell, *Thesis Guide - A Guide for Students in Computer Science and Information Systems*. 2008.

[6] I. Sommerville, *Software Engineering - 9th Edition*. 2010.

[7] R. Elmasri and S. B. Navathe, *Basics of Functional Dependencies and Normalization for Relational Databases*. 2010.

[8] Fathansyah, *Basis Data Edisi Revisi*. Bandung: Informatika, 2012.

[9] F. Frankel, *Visualizing Data*, vol. 92, no. 2. 2004.

[10] B. Shneiderman, “The eyes have it: a task by data type taxonomy for information\nvisualizations,” *Proc. 1996 IEEE Symp. Vis. Lang.*, pp. 336–343, 1996.

[11] A. Zoss, “LibGuides: Introduction to Data Visualization: Visualization Types.” [Online]. Available: http://guides.library.duke.edu/datavis/vis\_types. [Accessed: 14-Oct-2015].

[12] A. M. Glenberg and M. E. Andrzejewski, *Learning from data: an introduction to statistical reasoning*. 2008.