**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](#_Toc432605853)

[ABSTRACK i](#_Toc432605854)

[DAFTAR ISI ii](#_Toc432605855)

[DAFTAR GAMBAR iv](#_Toc432605856)

[DAFTAR TABEL v](#_Toc432605857)

[DAFTAR LAMPIRAN vi](#_Toc432605858)

[BAB I PENDAHULUAN 1](#_Toc432605859)

[I.1 Latar Belakang Masalah 1](#_Toc432605860)

[I.2 Perumusan Masalah 2](#_Toc432605861)

[I.3 Maksud dan Tujuan 2](#_Toc432605862)

[I.4 Batasan Masalah 2](#_Toc432605863)

[I.5 Metodologi Penelitian 2](#_Toc432605864)

[I.5.1 Metode Pengumpulan Data 3](#_Toc432605865)

[I.5.2 Metode Pengembangan Perangkat Lunak 3](#_Toc432605866)

[I.6 Sistematika Penulisan 4](#_Toc432605867)

[BAB II LANDASAN TEORI 6](#_Toc432605868)

[II.1 Konsep Data 6](#_Toc432605869)

[II.1.1 Data 6](#_Toc432605870)

[II.1.2 Basis Data 6](#_Toc432605871)

[II.2 Visualisasi Data 7](#_Toc432605872)

[II.2.1 Tipe Visualisasi 7](#_Toc432605873)

[II.2.2 Langkah Visualisasi Data 7](#_Toc432605874)

[II.3 Descriptive Statistics 7](#_Toc432605875)

[II.3.1 Distribusi Frekuensi dan *Percentile Rank* 7](#_Toc432605876)

[II.3.2 Ukuran Gejala Pusat dan Variasi 7](#_Toc432605877)

[II.3.3 Z Skor dan Distribusi Normal 9](#_Toc432605878)

[II.4 Inferential Statistics 9](#_Toc432605879)

[II.5 Algoritma 10](#_Toc432605880)

[DAFTAR PUSTAKA 1](#_Toc432605881)

# 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 terstruktur.
3. Sistem hanya menampilkan informasi visualisasi interaktif, tidak menerima data personal pengguna.

## Metodologi Penelitian

Metodologi penelitian yang digunakan dalam penelitian ini menggunakan pendekatan kuantitatif, metode kuantitatif memecahkan masalah dengan cara menguji hipotesis berdasarkan konsep yang dibangun dari teori-teori yang dikumpulkan [5]. Adapun metode 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*

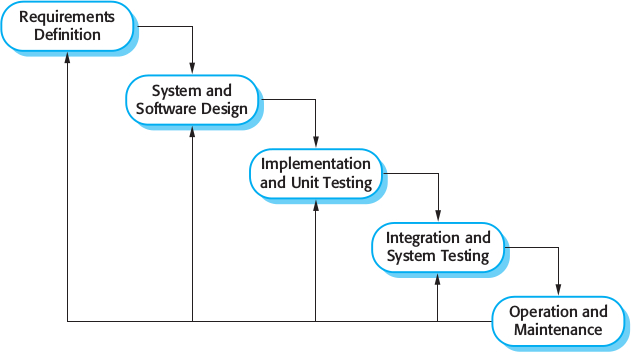
Tahap *Implementation and Unit testing* merupakan tahap dimana perancangan desain yang telah dibuat, diterjemahkan kedalam kode-kode unit program menggunakan bahasa pemrograman. Setelah itu sistem yang sudah jadi diuji apakah setiap unit mencapai spesifikasinya.

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 dikemudiah 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 unntuk 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 permasalah, merumuskan inti permasalahan, mencari solusi atas masalah yang muncul, melakukan identifikasi masalah tesebut, 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 system 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 Managemen 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 pentumpanan 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

### Tipe Visualisasi

### Langkah Visualisasi Data

## 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 [9]. *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 [9]. *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 [9].

### 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 [9].

#### 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 disteibusi. Formula median dapat dibedakan untuk jumlah distribusi genap dan ganjil.

…

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* [9].

Varian dari sebuah populasi dapat dihitung dengan cara [9]: (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

Adasdadad

## Inferential Statistics

Sdsfsafaf

## Algoritma

Leorm ipsum dolor sit amet

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