**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](#_Toc437981190)

[ABSTRACK ii](#_Toc437981191)

[DAFTAR ISI iii](#_Toc437981192)

[DAFTAR GAMBAR v](#_Toc437981193)

[DAFTAR TABEL vi](#_Toc437981194)

[DAFTAR LAMPIRAN x](#_Toc437981195)

[BAB I PENDAHULUAN 1](#_Toc437981196)

[I.1 Latar Belakang Masalah 1](#_Toc437981197)

[I.2 Perumusan Masalah 2](#_Toc437981198)

[I.3 Maksud dan Tujuan 2](#_Toc437981199)

[I.4 Batasan Masalah 2](#_Toc437981200)

[I.5 Metodologi Penelitian 2](#_Toc437981201)

[I.5.1 Metode Pengumpulan Data 3](#_Toc437981202)

[I.5.2 Metode Pengembangan Perangkat Lunak 3](#_Toc437981203)

[I.6 Sistematika Penulisan 4](#_Toc437981204)

[BAB II LANDASAN TEORI 6](#_Toc437981205)

[II.1 Konsep Data 6](#_Toc437981206)

[II.1.1 Data 6](#_Toc437981207)

[II.1.2 Basis Data 6](#_Toc437981208)

[II.2 Visualisasi Data 7](#_Toc437981209)

[II.2.1 Tipe-Tipe Visualisasi 7](#_Toc437981210)

[II.2.2 Proses Visualisasi Data 12](#_Toc437981211)

[II.3 Statistika Deskriptif 13](#_Toc437981212)

[II.3.1 Distribusi Frekuensi dan Persentil 14](#_Toc437981213)

[II.3.2 Ukuran Gejala Pusat dan Variansi 14](#_Toc437981214)

[II.3.3 *z* Skor dan Distribusi Normal 16](#_Toc437981215)

[II.4 Statistika Inferensi 17](#_Toc437981216)

[II.4.1 Probabilitas 17](#_Toc437981217)

[II.4.2 Distribusi Sampling 18](#_Toc437981218)

[II.5 Algoritma 18](#_Toc437981219)

[BAB III ANALISIS DAN PERANCANGAN DATA 20](#_Toc437981220)

[III.1. Analisis Visualisasi Data 20](#_Toc437981221)

[III.1.1 Analisis Masalah 20](#_Toc437981222)

[III.1.2 Analisis Sumber Data 20](#_Toc437981223)

[III.1.3 Analisis Segmentasi Pengguna Informasi 29](#_Toc437981224)

[III.1.4 Analisis Kebutuhan Informasi Berdasarkan Segmentasi Pengguna 29](#_Toc437981225)

[III.1.5 Analisis Algoritma 35](#_Toc437981226)

[III.1.6 Data Preprocessing 36](#_Toc437981227)

[III.1.7 Pemetaan Visualisasi Data 36](#_Toc437981228)

[III.2 Analisis Spesifikasi Kebutuhan Perangkat Lunak 36](#_Toc437981229)

[III.2.1 Analisis Kebutuhan Non Fungsional 36](#_Toc437981230)

[III.2.2 Analisis Kebutuhan Fungsional 36](#_Toc437981231)

[III.3 Perancangan Arsitektur 37](#_Toc437981232)

[III.3.1 Perancangan Struktur Menu 37](#_Toc437981233)

[III.3.2 Perancangan Antarmuka 37](#_Toc437981234)

[BAB IV IMPLEMENTASI DAN PENGUJIAN 38](#_Toc437981235)

[DAFTAR PUSTAKA 1](#_Toc437981236)

# DAFTAR GAMBAR

[Gambar I‑1 Model *Waterfall* [6] 4](#_Toc437981237)

[Gambar II‑1 Contoh Grafik Planar/2D [12] 9](#_Toc437981238)

[Gambar II‑2 Contoh Grafik Temporal [12] 9](#_Toc437981239)

[Gambar II‑3 Tree-map (Multi-dimensi) [12] 10](#_Toc437981240)

[Gambar II‑4 Contoh Radial-tree [12] 11](#_Toc437981241)

[Gambar II‑5 Contoh Node-diagram [12] 11](#_Toc437981242)

# DAFTAR TABEL

[Tabel III‑1 Data Laju Pertumbuhan Ekonomi Jakarta dan Nasional 26](#_Toc438252327)

[Tabel III‑2 Data Tingkat Inflasi Jakarta dan Nasional 26](#_Toc438252328)

[Tabel III‑3 Data Komponen Inflasi Jakarta 26](#_Toc438252329)

[Tabel III‑4 Data Ekspor Impor DKI Jakarta 26](#_Toc438252330)

[Tabel III‑5 Data Struktur Ekonomi Jakarta Berdasarkan Sektor 27](#_Toc438252331)

[Tabel III‑6 Data Ikhtisar Statistik Antar Kerja DKI Jakarta 27](#_Toc438252332)

[Tabel III‑7 Data Upah Minimum Provinsi dan Inflasi 27](#_Toc438252333)

[Tabel III‑8 Data Laporan Realisasi Anggaran 28](#_Toc438252334)

[Tabel III‑9 Data Laporan Neraca 28](#_Toc438252335)

[Tabel III‑10 Data Anggraran Belanja DKI Jakarta Per Kegiatan 28](#_Toc438252336)

[Tabel III‑11 Data APBD Perubahan DKI Jakarta 2014 29](#_Toc438252337)

[Tabel III‑12 Data APBD Penetapan DKI Jakarta 2014 29](#_Toc438252338)

[Tabel III‑13 Data Pendapatan Perkapitra Jakarta dan Nasional 30](#_Toc438252339)

[Tabel III‑14 Data Anggaran Pendapatan Belanja Daerah 30](#_Toc438252340)

[Tabel III‑15 Data Anggaran Belanja Langsug dan Tidak Langsung 30](#_Toc438252341)

[Tabel III‑16 Data Anggaran dan Realisasi Pendapatan Daerah menurut Sumber dan Jenis Pengeluaran 31](#_Toc438252342)

[Tabel III‑17 Data Jumlah Pendapatan DKI Jakarta dari Pajak 31](#_Toc438252343)

[Tabel III‑18 Data Anggaran dan Realisasi Penerimaan dan Pengeluaran 31](#_Toc438252344)

[Tabel III‑19 Data Realisasi APBD DKI Jakarta per SKPD 32](#_Toc438252345)

[Tabel III‑20 Data APBD DKI Jakarta Berdasarkan Kegiatan 32](#_Toc438252346)

[Tabel III‑21 Data Realisasi APBD DKI Jakarta per Kegiatan 33](#_Toc438252347)

[Tabel III‑22 Data Ringkasan Anggaran dan Realisasi, Pendapatan, Belanja, dan Pembiayaan 33](#_Toc438252348)

[Tabel III‑23 Data Pertumbuhan Ekonomi Jakarta dan Nasional 37](#_Toc438252349)

[Tabel III‑24 Data Pendapatan Perkapita Jakarta dan Nasional 38](#_Toc438252350)

[Tabel III‑25 Data Inflasi Jakarta dan Nasional 38](#_Toc438252351)

# 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 tempat penyimpanan *“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 tempat penyimpanan data Pemerintah Provinsi DKI Jakarta masih sulit dibaca dan dimengerti karena format data yang terbilang asing 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 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. 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 pada grup facebook UKM atau *Entrepreneur* melalui fasilitas google form.

1. Pencarian Dokumen

Dokumen yang digunakan adalah dokumen yang bersumber dari tempat penyimpanan *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*

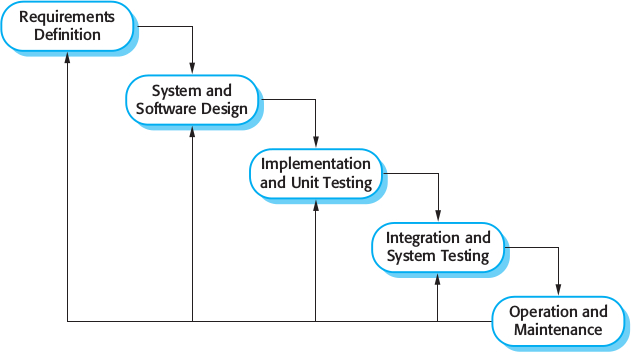
Tahap *Implementation and Unit testing* merupakan tahap dimana perancangan desain yang telah dibuat, diterjemahkan ke dalam kode-kode unit program menggunakan bahasa pemrograman. Setelah itu sistem yang sudah jadi diuji apakah setiap unit mencapai spesifikasinya.

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 fungsionalitas 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 didesain dan dibangun dengan tujuan tertentu [7]. Perangkat lunak yang membantu pengguna dalam mengelola dan membuat basis data disebut dengan DBMS (*Data Base Management System*) [8]. Tujuan utama dari DBMS adalah menyediakan cara untuk menyimpan dan mengambil informasi dari basis data secara lebih mudah dan efisien.

Sistem basis data didesain untuk mengelola informasi yang besar [9]. Pengelolaan data termasuk mendefinisikan struktur informasi penyimpanan dan penyediaan mekanisme untuk melakukan manipulasi informasi. Selain itu, sistem basis data juga harus bisa memastikan bahwa informasi disimpan dengan aman, baik dari *system crash* atau akses yang tidak diizinkan. Jika data memang dikelola oleh beberapa pengguna, sistem harus menghindari sebisa mungkin hasil yang anomali.

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 (*data-definition language*) dan DML (*data-manipulation language*) [8].

1. DDL (*Data Definition Language*)

DDL digunakan untuk membuat skema dan memberikan properti tambahan data, dengan kata lain 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 memungkinkan pengguna DBMS mengakses atau melakukan manipulasi data yang diorganisasikan dengan model data yang sesuai. Tipe akses tersebut yaitu:

* Mengambil (*retrieve*) informasi dari basis data.
* Memasukan (*insert*) informasi baru ke basis data.
* Menghapus (*delete*) infromasi dari basis data.
* Modifikasi (*update*) informasi yang terdapat dari basis data.

### Basis Data NoSQL

NoSQL diinterpretasikan dengan *“not only SQL”* (bukan SQL saja) DBMS yang tidak memakai model relasi; NoSQL tidak dibangun di atas tabel, sehingga tidak bisa memakai query SQL untuk melakukan manipulasi data. Keunggulan NoSQL dibandingkan basis data relasi ada pada skabilitas [10]. Selain skalabilitas, NoSQL memiliki keunggulan sebagai berikut [10]:

* *Scemaless data representation*: Hampir semua model NoSQL merepresentasikan *schemalees*. Administrator basis data tidak perlu memikirkan bagaimana struktur dan model basis data yang berubah setiap waktu.
* *Development time*: Query yang komplek pada basis relasi yang sudah besar, akan membuat proses penyajian data menjadi lambat (misalnya JOIN tabel pada basis data yang tersebar di banyak server). Pada NoSQL hal demikian tidak terjadi.
* *Plan ahead for scalability*: Aplikasi yang memakai model NoSQL akan sangat elastis, NoSQL dapat menangani masalah lonjakan sumber daya.

## 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 secara jelas, melihat pola, dan menunjukan informasi dari berbagai dimensi. Visualisasi yang efektif membantu pengguna dalam melakukan analisis, mempermudah menyimpulkan data, dan membuat data yang kompleks menjadi lebih mudah untuk dimengerti [11].

Visualisasi data tidak semerta-merta mengubah data menjadi grafik visual, visualisasi data memerlukan perencanaan. Setiap jenis data memiliki teknik visualisasi berdasarkan kebutuhannya. Berdasarkan tingkat kompleksitas data, untuk menghasilkan solusi yang berharga diperlukan dari berbagai disiplin ilmu, yaitu: statistika, *data mining*, desain grafis, dan visualisasi informasi [11].

### *Visual Encoding*

Visual encoding adalah sebuah cara memtakan data ke dalam struktur visual. Desainer membantu pembaca memahami makna data lebih cepat dengan memakai beberapa prinsip *visual encoding*. Terdapat dua tipe prinsip *visual encoding*: planar dan retina. Manusia sangat peka terhadap variabel retina [12].

Terdapat tiga tipe data dasar dalam *visual encoding*. Berikut adalah tiga tipe dasar *visual encoding* [12]:

1. Kuantitatif: Tipe data ini dibedakan berdasarkan objek yang bisa diukur. Contoh: Nilai ujian, durasi jam dalam sehari, jumlah hari dalam setahun, dll.
2. Kualitatif: Tipe data ini dibedakan berdasarkan objek yang bisa dibandingkan dan tidak memiliki ukuran yang pasti. Contoh: Baik, buruk, jelek, tidak yakin, dll.
3. *Categorical*: Tipe data ini dibedakan berdasarkan pengelompokan objek. Contoh:
   1. Buah: apel, jeruk, mangga, dll;
   2. Negara: Singapura, Indonesia, Jepang, Rusia, dll;
   3. dll.

### Tipe-Tipe Visualisasi

Tipe visualisasi data dibedakan berdasarkan tujuan yang hendak ingin dicapai. Terdapat tujuh hal yang harus dipenuhi dalam melakukan abstraksi tingkat tinggi (*high-level abstraction*), semakin banyak hal yang disembunyikan, semakin banyak juga langkah-langkah yang harus dipenuhi, langkah tersebut yaitu [13]:

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.

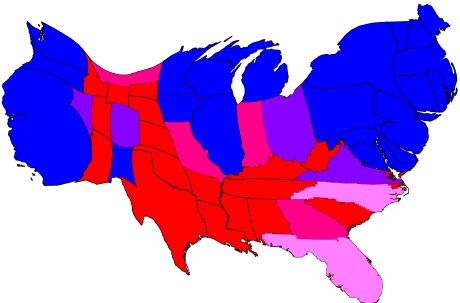
Berdasarkan taksonominya, grafik visual dibedakan menjadi [13]: 1D/Linear; 2D/Planar; 3D/Volumetric; Temporal; Multidimensional; Tree/Hierarchical; dan Network.

#### 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).



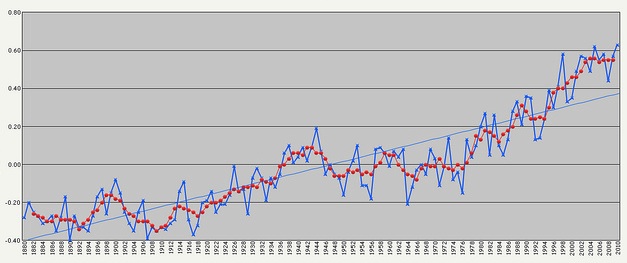
Gambar II‑1 Contoh Grafik Planar/2D [14]

#### 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

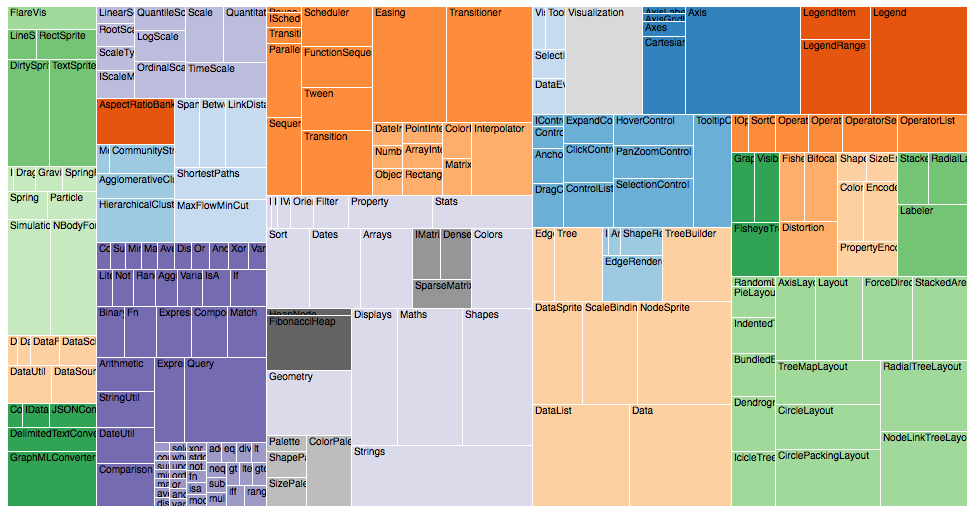
Grafik temporal adalah grafik yang berhubungan dengan waktu (*time lines*). Grafik ini menggambarkan persentasi historikal dari data 1-dimensi. Yang membedakan, grafik temporal memiliki *item* dengan waktu awal dan waktu akhir, atau periode tertentu.



Gambar II‑2 Contoh Grafik Temporal [14]

#### Multi-dimensional

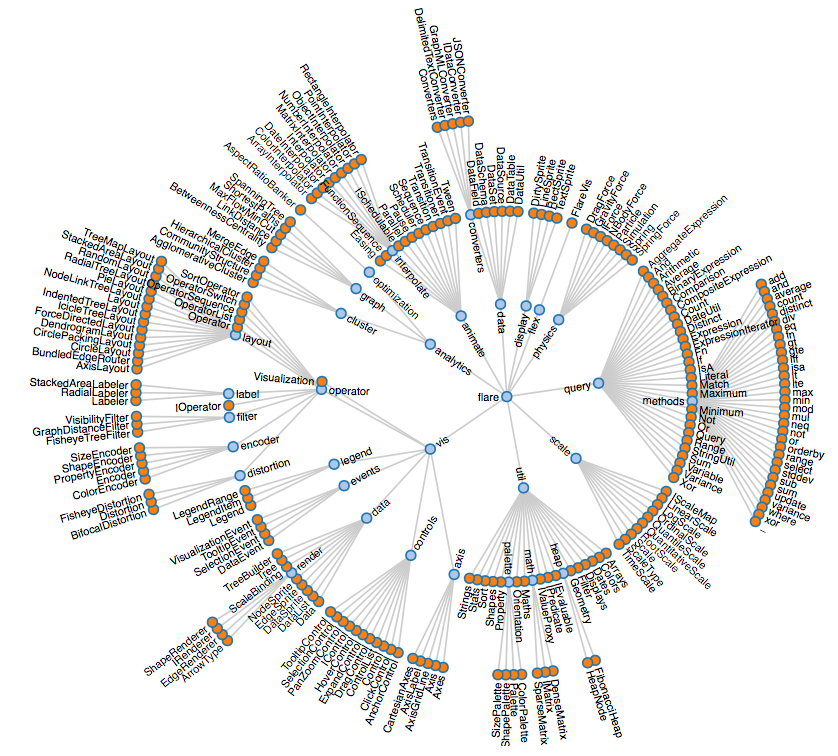
Grafik temporal didalamnya termasuk grafik-grafik yang dihasilkan dari manipulasi data dari disiplin ilmu statistika. Antarmuka representasi milti-dimensional adalah grafik 2-dimensi. Grafik multi-dimensi termasuk didalamnya grafik *pie*, histogram, *tag cloud*, *bubble cloud*, bar, *tree-map*, *scatter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



Gambar II‑3 Tree-map (Multi-dimensi) [14]

#### Tree/Hierarchical

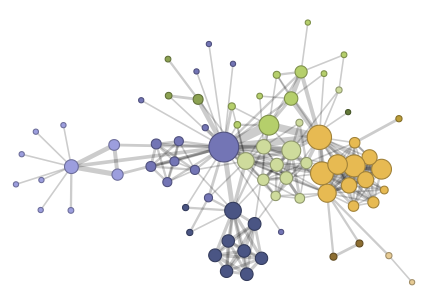
*Tree* adalah grafik herarkikal dari item-item yang memiliki hubungan satu dengan lainnya, atau yang memiliki induk (kecuali *root*). Setiap *item* antara induk dan anak bisa memiliki banyak atribut. Grafik *tree* termasuk didalamnya grafik *tree, dendorogram, radial-tree, hyperbolic-tree, tree-map*,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



Gambar II‑4 Contoh Radial-tree [14]

#### 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‑5 Contoh Node-diagram [14]

### 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 baik tentu terfokus, memberikan jawaban yang jelas, dan tidak terlalu detail [11]. Banyak metode dan cara untuk membuat visualisasi yang baik, akan tetapi pada prinsip dasarnya, visualisasi yang bagus memenuhi aspek: *overview*, *zoom* dan *filtering*, selanjutnya *details-on-demand* [13].

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 [11]:

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, seperti 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.

## Statistika Deskriptif

Statistika deskriptif (*descriptive statistics*) adalah suatu metode untuk mengorganisasikan dan melihat gambaran umum dari pengukuran sampel dan populasi. Statistika deskriptifseperti namanya biasa dipakai untuk membaca data, seringkali digunakan pada sebuah sampel populasi karena sulitnya untuk mendapatkan seluruh data populasi [15]. Bila perhitungan diperoleh dari sampel yang diambil dari populasi, maka hasil tersebut disebut statistika. Sementara bila perhitungan ditarik dari seluruh populasi, maka hasilnya disebut parameter.

Statistika deskriptif 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 statistika deskriptif, distribusi frekuensi dan peresentil, titik tengah dan variabilitas, dan *z* skor dan distribusi normal.

### Distribusi Frekuensi dan Persentil

Distribusi frekuensi adalah tabulasi jumlah kemunculan dari setiap nilai. Distribusi frekuensi mengelola dan memberikan gambaran umum data, yaitu dengan cara menampilkan karakteristik terbanyak [15]. Persentil adalah persentasi skor nilai yang didapatkan dari perhitungan distribusi di bawah skor nilai. Persentil dan persen tidaklah sama. Skor nilai sendiri sering disebut sebagai persentil (*P*th), adalah skor nilai *P*% dari perhitungan di bawah distribusi [15].

### Ukuran Gejala Pusat dan Variansi

Ukuran gejala pusat distribusi adalah nilai yang dekat dengan titik tengah dari distribusi. Biasanya digunakan sebagai representasi nilai distribusi keseluruhan [15].

#### 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 [1], 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, mean, median, dan modus (*mode*). Nilai pada mean, 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* (miu).

|  |  |
| --- | --- |
| *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.

#### Varian dan Standar Deviasi

Variabilitas adalah pengukuran beda nilai antara satu nilai distribusi dengan nilai distribusi pusat atau seberapa besar penyimpangan dengan nilai pusatnya. Pengukuran yang lebih menggambarkan variabilitas adalah varian, varian menghitung variabilitas berdasarkan nilai beda dengan *µ* [15].

Varian dari sebuah populasi dapat dihitung dengan cara [15]: (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:

Varian dari populasi atau sampel tidak dapat dinyatakan dalam stauan ukur, karena alasan tersebut ukuran varian yang sering digunakan adalah standar deviasi. Secara matematis, standar deviasi didefinisikan sebagai akar kuadrat dari varian.

*(parameter)*

*(statistik)*

### *z* Skor dan Distribusi Normal

*z* skor atau disebut juga dengan skor standar, sama seperti persentil, *z* skor juga menghitung posisi relatif dari mean. Index *z* skor adalah posisi distribusi relatif terhadap distribusi lain. *z* skor sangat berguna pada distribusi normal [15].

*z* skor adalah skor yang sudah dilakukan standarisasi dengan menjadikan 0 sebagai mean dari distribusi dan 1 sebagai standar deviasi [15]. Formula untuk menghitung *z* skor adalah sebagai berikut:

Jarak skor dari mean berdasarkan *z* skor adalah: + dan +2 di atas nilai mean () atau : - dan -2 di bawah nilai mean ().

## Statistika Inferensi

Tahap pertama memahami data adalah memahami sampel, selanjutnya, sampel yang diambil seringkali digunakan untuk menyimpulkan kondisi populasi sesungguhnya dimana sampel tersebut diambil. Statistika inferensi menggunakan sembarang sampel dari populasi yang diuji untuk membuat kesimpulan atau menaksir terhadap karakteristik populasi [15]. Statistika inferensi 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 danpengujian hipotesis.

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

### Probabilitas

Probabilitas adalah kemungkinan yang dapat terjadi dalam suatu peristiwa (*event*) tertentu. Secara matematis, probabilitas diartikan sebagai banyaknya pristiwa yang dimaksud dibagi dengan seluruh pristiwa yang mungkin [15]. Formula probabilitas yaitu:

dengan:

*P*(*A*) = probabilitas terjadinya peristiwa *A*

*n*(*A*) = jumlah peristiwa *A*

*n*(*S*) = jumlah peristiwa yang mungkin

### Distribusi Sampling

Keputusan diambil terhadap informasi yang diambil dari populasi. Informasi yang lengkap diperoleh dari seluruh populasi. Hanya saja, untuk mendapatkan seluruh data dari populasi hampir tidak mungkin dilakukan secara singkat. Sampling adalah metode yang digunakan untuk mengestimasi parameter-parameter populasi [15].

Standar deviasi dari distribusi sampling disebut dengan standar error statistik (*SE*). Perbedaan standar deviasi dan standar error adalah pada nilai, sementara *standard error* berkaitan dengan hasil perhitungan.

Perbedaan hasil yang diperoleh dari sampel statistik dan hasil yang diperoleh dari parameter populasi disebut sampling error. Sampling error dipakai untuk menunjukan ketepatan populasi yang didasarkan pada perhitungan terhadap sampel di bawah distribusi normal. Semakin kecil sampling error, semakin besar ketepatan penaksiran.

#### Pendugaan Titik

Cara penaksiran skor tunggal parameter populasi mean (*µ*) dapat dilakukan dengan mengambil mean terhadap sampel acak populasi (*M*). Mean sampel acak tersebut disebut pendugaan titik (*point estimator --* ) [15].

#### Pendugaan Interval

Pendugaan interval (*interval estimation*) menunjukan interval berapa suatu parameter populasi akan berada pada dua nilai (dibatasi oleh dua nilai) karena adanya *treatment*, nilai tersebut biasanya disebut batas bawah dan batas atas [15].

#### *Significance of Alpha Level (α)*

Alpha level (*α*) adalah simbol yang menunjukan sebarapa besar suatu kejadian dalam sampel terjadi akibat pengambilan acak (*likely*) atau bukan acak (*unlikely*). Artinya kejadian yang muncul dari sampel apakah terjadi secara kebetulan (*random*) atau bukan kebetulan [16]. *Significance of Alpha level (α)* dibagi menjadi tiga, yaitu:

1. *α* = 0.05 (5%)
2. *α =* 0.01 (1%)
3. *α =* 0.001 (0.1%)

#### Nilai Statistik *t*

Nilai statistik *t* menggambarkan seberapa jauh beda rata-rata sampel antara satu sampel dengan sampel lainnya atau populasi. Nilai Statistik *t* didapatkan dari pembagian antara beda mean populasi dengan sampel dengan *standard error* [15]*.*

## Korelasi

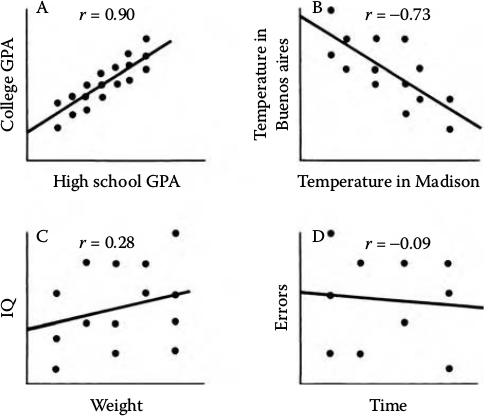
Korelasi digunakan untuk mengukur hubungan antara dua sampel, yaitu ketika pengukuran setiap variabel sampel dengan variabel lainnya harus dilibatkan. Analisis korelasi digunakan untuk membuat kesimpulan terhadap populasi dari sampel yang diukur, analisis ini bisa dilakukan hanya ketika suatu sampel yang dependen [15].

### Korelasi *Pearson* *r*

Korelasi *Pearson r* atau yang sering disebut *r* saja, adalah pengukuran terhadap kekuatan garis linear antara dua variabel. Pengukuran *r* bisa diaplikasikan pada sampel atau populasi, simbol *r* dugunakan untuk mengukur statistik, sedangkan untuk parameter menggunakan simbol *Rho* (*ρ*) [15]. Baik *r* atau *ρ* memiliki properi yang sama, yaitu [15]:

1. Nilai korelasi selalu pada rentang -1 dan 1m inklusif.
2. Nilai korelasi selalu menunjukan hubungan dengan kemiringan/gradien garis regresi; garis regresi dengan kemiringan positif memiliki koefisien korelasi positif, dan sebaliknya. Garis regresi dengan gradien 0, memiliki koefisien korelasi 0.0.
3. Kekuatan relasi garis linear diindikasikan dengan nilai absolut dari korelasi. Relasi garis linear yang sempurna ditandai dengan *r* atau *ρ* yang bernilai -1 atau 1; tidak ada hubungan garis linear *r* atau *ρ* bernilai 0.

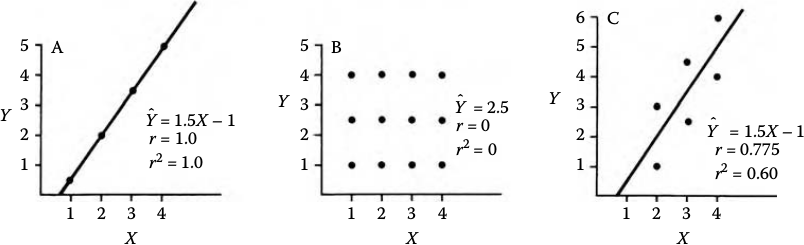
Kekuatan relasi garis linear ditandai dengan *scatter plots* yang dekat dengan garis regresi (*Sy.x* kecil). Ilustrasi ini digambarkan pada Gambar II‑6.



Gambar II‑6 Ilustrasi Korelasi dan *Scatter plots* [15]

### Interpretasi *r2*

Kelemahan dari *r* atau *ρ* adalah tidak memiliki interpretasi langsung terhadap varian. *r2* atau *ρ2* dapat menghitung proporsi *y* varian terhadap *x* varian secara jelas [15]. Interpretasi dari *r2* digambarkan pada Gambar II‑7.



Gambar II‑7 Ilustrasi Korelasi *r2*

Secara matematis, r2 dapat dihitung dengan mengurangi varian *x* terhadap *x* dengan total varian dari *y* dengan 1 [17].

### Faktor-Faktor yang Memengaruhi Nilai *r*

Terdapat beberapa faktor yang harus diperhatikan ketika akan menghitung hubungan antara *x* dan *y*. Jika diabaikan, faktor-faktor tersebut akan memberikan masalah terhadap regresi [15]. Faktor-faktor tersebut yaitu [15]:

1. *Nonlinear relationship*. Korelasi linear hanya menghitung derajat hubungan kelineran. Jika *x* dan *y* berhubgan tetapi tidak linear, *r* dapat memberikan nilai kosong (meskipun terdapat hubungan antara dua variabel).
2. *Restricted range*. Pengecualian baik pada range *x* atau *y*, akan mengurangi nilai *r.*
3. *Extreme score*. Satu nilai yang sangat tinggi dapat menghasilkan bukti pada saat perhitungan korelasi ketika tidak ditemukan pada skor mayoritas.
4. *Combining griups.* Dapat terjadi tidak adanya hubungan antara grup, tetapi menggabungkan beberapa grup dapat menimbulkan kebiasan dalam korelasi linear.

### Pendugaan Interval *Rho (ρ)*

Nilai statistik *r* dapat digunakan untuk membuat berbagai macam kesimpulan tentang *ρ*. Salah satu manfaat *r* adalah pendugaan interval terhadap *ρ*. Dua nilai *r* dapat dihitung untuk menduga nilai *ρ*. Sama seperti pendugaan interval pada , dimana diperlukan nilai standar yaitu *z* dan *α* level*.* Pada pendugaan interval *ρ* juga diperlukan nilai statistik *t* dan *α* level [15]. Nilai statistik t untuk menduga interval *ρ* dapat dihitung dengan formula:

*df* adalah kependekan dari *dgree of freedom* (derajat kebebasan). Dimana *df* dapat dihitung dengan mengurangi setiap kelompok sampel dengan 1. Artinya, jika terdapat *n* sampel, maka *df* adalah *n*-(banyak sampel). Sehingga, untuk regresi linear yang memiliki sampel *x* dan sampel *y,* maka *df* adalan 2(*n*-1). Selanjutnya pada *t* statistik dinotasikan *N*-2 [15].

Selanjutnya interval *ρ* dapat dihitung dengan formula [18]:

adalah simpangan baku dari *r.* dapat dihitung dengan formula:

### Pengambilan Keputusan

Untuk membuat keputusan, apakah garis regresi yang telah dibuat mewakili kondisi sebenarnya (populasi) atau belum, maka diperlukan nilai dari *α* level dan titik kritis *t*. Jika *r* jatuh pada interval *ρ* dengan titik kritis maka nilai peramalan dapat muncul bukan karena acak (kebetulan) [15].

## Regresi Linear

Regresi linear merupakan salah satu pembelajaran mesin *supervised learning*, yaitu pembelajaran mesin yang memprediksi nilai yang nyata [19]. Regresi linear memprediksi nilai dengan membuat model. Model adalah represtasi dari *state*, proses, atau sistem untuk mengambil sebuah kesimpulan [19]. Regresi linear menjelaskan hubungan matematis antara dua variable. Regresi linear memakai variabel independen *x* untuk menjelaskan variabel dependen *y* atau sering disebut juga *outcome variable* [20].

Model regresi linear yang terbentuk untuk sampel adalah , sedangkan untuk populasi adalah . *b* adalah tingkat kemiringan untuk sampel, sedangkan *β*1  adalah tingkat kemiringan (*slop*) untuk populasi. Untuk mencari tingkat kemiringan, dilakukan dengan cara mengalikan *r* dengan pembagian *SE*y dengan *SE****x*** [16]*.*

Model regresi linear ini kemudian diplot melewati titik (,) [16].

# ANALISIS DAN PERANCANGAN DATA

## Analisis Visualisasi Data

Analisis visualisasi data yang dilakukan yaitu melakukan proses analisis masalah, analisis sumber data, analisis segmentasi pengguna informasi, analisis kebutuhan informasi berdasarkan segmentasi pengguna, analisis algoritma, dan data preprosessing, dan pemetaan visualisasi data.

### Analisis Masalah

Berikut merupakan analisis permasalahan yang coba diselesaikan dalam penelitian visualisasi data ini:

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. Data yang digunakan dalam penelitian ini adalah dataset perekonomian dan keuangan daerah.

#### Dataset Perekonomian

Pada dataset perekonomian, terdapat enam data yang bisa digunakan secara bebas. Dataset perekonomian meliputi dataset tentang laju pertumbuhan ekonomi, dataset tentang inflasi, dataset ekspor impor, dataset struktur ekonomi, dll. Tabel-tabel yang terdapat pada dataset perekonomian adalah sebagai berikut:

1. Tabel Data Laju Pertumbuhan Ekonomi Jakarta dan Nasional

Tabel ini berisi data mengenai besar laju pertumbuhan ekonomi Jakarta dan Nasional tahun 2006-2012. Berikut adalah penjelasan atribut pada tabel ini:

Tabel III‑1 Data Laju Pertumbuhan Ekonomi Jakarta dan Nasional

| Atribut | Keterangan |
| --- | --- |
| Tahun | Tahun |
| persen\_tumbuh\_jakarta | Pertumbuhan ekonomi Jakarta (dalam persen) |
| Persen\_tumbuh\_nasional | Pertumbuhan ekonomi Nasional (dalam persen) |

1. Tabel Data Tingkat Inflasi Jakarta dan Nasional

Tabel ini berisi data tantang tingkat inflasi Jakarta dan Nasional tahun 2006-2012. Berikut adalah penjelasan atribut pada tabel ini:

Tabel III‑2 Data Tingkat Inflasi Jakarta dan Nasional

|  |  |
| --- | --- |
| Atribut | Keterangan |
| Tahun | Tahun |
| inflasi\_jakarta | Inflasi Jakarta (dalam persen) |
| inflasi\_nasional | Inflasi Nasional (dalam persen) |

1. Tabel Data Komponen Inflasi Jakarta

Tabel ini berisi data mengenai faktor-faktor dan komponen inflasi di DKI Jakarta per bulan tahun 2012. Berikut adalah penjelasan atribut pada tabel ini:

Tabel III‑3 Data Komponen Inflasi Jakarta

| Atribut | Keterangan |
| --- | --- |
| komponen\_inflasi | Komponen inflasi |
| Bulan | Bulan |
| bersen\_inflasi | Persen inflasi dari komponen tersebut |

1. Tabel Data Ekspor dan Impor DKI Jakarta

Tabel ini berisi data besaran dan nilai ekspor DKI Jakarta pada tahun 2006- 2012. Berikut adalah penjelasan atribut pada tabel ini:

Tabel III‑4 Data Ekspor Impor DKI Jakarta

| Atribut | Keterangan |
| --- | --- |
| Tahun | Tahun |
| ekspor\_melalui\_jakarta | Nilai ekspor melalui Jakarta (dalam juta USD) |
| ekspor\_produk\_jakarta | Nilai ekspor produk Jakarta (dalam juta USD) |
| impor\_melalui\_jakarta | Nilai Impor melalui Jakarta (dalam juta USD) |

1. Tabel Data Struktur Ekonomi Jakarta Berdasarkan Sektor

Tabel ini berisi data struktur ekonomi Jakarta berdasarkan persentase kontribusi per jenis kegiatan tahun 2006-2012. Berikut adalah penjelasan atribut pada tabel ini:

Tabel III‑5 Data Struktur Ekonomi Jakarta Berdasarkan Sektor

| Atribut | Keterangan |
| --- | --- |
| tahun | Tahun |
| jenis\_sektor | Jenis sector ekonomi (Primer, Sekunder, dan Tersier) |
| persen\_kontribusi | Persentase kontribusi dari sektor ekonomi tersebut |

1. Tabel Data Ikhtisar Statistik Antar Kerja DKI Jakarta

Tabel ini berisi data iktisar statistik antar kerja di Provinsi DKI Jakarta pada tahun 2009-2012. Berikut adalah penjelasan atribut pada tabel ini:

Tabel III‑6 Data Ikhtisar Statistik Antar Kerja DKI Jakarta

| Atribut | Keterangan |
| --- | --- |
| tahun | Tahun |
| indikator | Indikator |
| rincian\_indikator | Rincian dari indicator |
| jumlah | Jumlah |

1. Tabel Data Upah Minimum Provinsi dan Inflasi

Tabel ini berisi data upah minimum Provinsi dan Inflasi di DKI Jakarta dari tahun 1997-2013. Berikut adalah penjelasan atribut pada tabel ini

Tabel III‑7 Data Upah Minimum Provinsi dan Inflasi

| Atribut | Keterangan |
| --- | --- |
| tahun | Tahun |
| ump | Upah Minimum Provinsi |
| kenaikan\_ump | Kenaikan Upah Minimum Provinsi |
| inflasi | Besarnya Inflasi |

#### Dataset Keuangan Daerah

Pada dataset keuangan daerah, terdapat 15 data yang bisa digunakan secara bebas. Dataset keuangan daerah meliputi dataset laporan realisasi anggaran, laporan neraca, anggaran belanja daerah, APBD perubahan, pendapatan perkapitra, jumlah pendapatan, dll. Tabel-tabel yang terdapat pada dataset perdagangan adalah sebagai berikut:

1. Tabel Data Laporan Realisasi Anggaran

Tabel ini berisi mengenai laporan realisasi anggaran DKI Jakarta pada tahun 2009-2012. SILPA padah tahun 2009 sebesar Rp3.756.718.028.446,33 termasuk Sisa Kas pada Bendahara Pengeluaran Kota Administrasi Jakarta Barat yang belum disetor ke Kas Daerah sebesar Rp.8.205.881.148,00 yang akan di reklas setelah ada keputusan. Berikut adalah penjelasan atribut pada tabel ini:

Tabel III‑8 Data Laporan Realisasi Anggaran

| Atribiut | Keterangan |
| --- | --- |
| nomor\_urut | Nomor urut |
| tahun | Tahun anggaran |
| pos\_anggaran | Pos anggaran (belanja, pendapatan, pembayaran) |
| sumber\_pos\_anggaran | Sumber dari pos anggaran |
| penjelasan\_pos\_anggaran | Penjelasan mengenai pos anggaran |
| catatan | Catatan |
| anggaran\_setelah\_perubahan | Nilai pos anggaran |
| realisasi | Realisasi pos anggaran |

1. Tabel Data Laporan Neraca

Tabel ini berisi menganai laporan neraca pemerintahan DKI Jakarta tahun 2008-2012. Berikut adalah penjelasan atribut pada tabel ini:

Tabel III‑9 Data Laporan Neraca

| Atribut | Keterangan |
| --- | --- |
| pos\_neraca | Pos neraca (aset, kewajiban, ekulitas) |
| sumber\_pos\_neraca | Sumber dari pos neraca |
| penjelasan\_pos\_neraca | Penjelasan mengenai pos neraca |
| catatan | Catatan |
| tanggal | Tanggal per pelaporan (sudah teraudit) |
| nilai | Nilai dari pos neraca terkait |

1. Tabel Data Anggaran Belanja DKI Jakarta per Kegiatan

Tabel ini berisi mengenai nilai anggaran untuk belanja langsung Pemda DKI Jakarta per kegiatan tahun 2013-2014. Berikut adalah penjelasan atribut pada tabel ini:

Tabel III‑10 Data Anggraran Belanja DKI Jakarta Per Kegiatan

| Atribut | Keterangan |
| --- | --- |
| komisi | Nomor komisi |
| skpdid | ID SKPD |
| skpd\_kode | Kode SKPD |
| belanja | Jenis Belanja |
| nama\_skpd | Nama SKPD |
| id\_kegiatan | ID kegiatan |
| kegiatan | Nnama kegiatan |
| nilai\_anggaran | Nilai anggaran (dalam rupiah) |

1. Tabel Data APBD Perubahan

Tabel ini berisi mengenai APBD perubahan DKI Jakarta pada tahun 2014 berdsarkan kegiatan (tanpa realisasi penyerapan). Berikut adalah penjelasan atribut pada tabel ini:

Tabel III‑11 Data APBD Perubahan DKI Jakarta 2014

| Atribut | Keterangan |
| --- | --- |
| lokasi\_id | Lokasi id kegiatan (provinsi) |
| Komisi | Komisi DPRD |
| skpd\_kode | Kode SKPD |
| skpd\_nama | Nama SKPD |
| urusan\_kode | Kode urusan |
| urusan\_nama | Nama urusan |
| program\_kode | Kode program |
| program\_teks | Nama program |
| kegiatan | Nama kegiatan |
| Anggaran | Nilai anggaran kegiatan |
| Vol | Volume kegiaran |
| Satuan | Satuan kegiatan |

1. Tabel Data APBD Penetapan DKI Jakarta Tahun 2014

Tabel ini berisi mengenai APBD penetapan DKI Jakarta tahun 2014 berdasarkan kegiatan (tanpa realisasi penyerapan). Berikut adalah penjelasan atribut pada tabel ini:

Tabel III‑12 Data APBD Penetapan DKI Jakarta 2014

| Atribut | Keterangan |
| --- | --- |
| lokasi\_id | Lokasi id kegiatan provinsi |
| komisi | Komisi DPRD |
| skpd\_kode | Kode SKPD |
| skpd\_nama | Nama SKPD |
| urusan\_kode | Kode urusan |
| urusan\_nama | Nama urusan |
| program\_kode | Kode program |
| program\_teks | Nama program |
| Kegiatan | Nama Kegiatan |
| Anggaran | Nilai anggaran kegiatan |
| Vol | Volume kegiatan |
| Satuan | Satuan kegiatan |

1. Tabel Data Pendapatan Perkapitra Jakarta dan Nasional

Tabel ini berisi besaran pendapatan per kapitra DKI Jakarta dengan Nasional tahun 2006-2012. Berikut adalah penjelasan atribut pada tabel ini:

Tabel III‑13 Data Pendapatan Perkapitra Jakarta dan Nasional

| Atribut | Keterangan |
| --- | --- |
| tahun | Tahun |
| perkapitra\_jakarta | Pendapatan per kapitra DKI Jakarta (dalam juta rupiah) |
| perkapitra\_nasional | Pendapatan per kapitra Nasional (dalam juta rupiah) |

1. Tabel Data Anggaran Pendapatan Belanja DKI Jakarta

Tabel ini berisi mengenai Anggaran Pendapatan Belanja Daerah (APBD) DKI Jakarta tahun 2014. Berikut adalah penjelasan atribut pada tabel ini:

Tabel III‑14 Data Anggaran Pendapatan Belanja Daerah

| Atribut | Keterangan |
| --- | --- |
| nomor\_urut | Nomor urut |
| pos\_anggaran | Pos anggaran (pendapatan, belanja, atau pembayaran) |
| sumber\_pos\_anggaran | Sumber dari pos anggaran |
| penjelasan\_pos\_anggaran | Penjelasan dari pos anggaran |
| nilai\_anggaran | Nilai anggaran yang berasal dari pos tersebut |

1. Tabel Data Anggaran Belanja Langsung dan Tidak Langsung

Tabel ini berisi mengenai nilai anggaran untuk belanja langsung dan tidak langsung dari Pemda DKI Jakarta untuk tahun 2013. Berikut adalah penjelasan atribut pada tabel ini:

Tabel III‑15 Data Anggaran Belanja Langsug dan Tidak Langsung

| Atribut | Keterangan |
| --- | --- |
| id\_skpd | ID dari SKPD |
| wilayah | Nomor wilayah |
| komisi | Nomor komisi |
| bidang | Nomor bidang |
| kode\_skpd | Kode SKPD |
| nama\_skpd | Nama SKPD |
| jenis\_belanja | Jenis belanja (langsung atau tidak langsung) |
| keterangan\_belanja | Keterangan mengenai belanja |
| nilai\_anggaran | Nilai anggaran pos belanja tersebut (dalam rupiah) |

1. Tabel Data Anggaran Realisasi dan Penerimaan Pengeluaran Daerah Menurut Sumber dan Jenis Kegiatan

Tabel ini berisi data anggaran dan realisasi pendapatan daerah menurut sumber dan jenis pengeluaran tahun 2010-2012. Berikut adalah penjelasan atribut pada tabel ini:

Tabel III‑16 Data Anggaran dan Realisasi Pendapatan  
Daerah menurut Sumber dan Jenis Pengeluaran

| Atribut | Keterangan |
| --- | --- |
| tahun | Tahun |
| akun | Keterangan anggaran |
| kelompok | Kelompok penerimaan dan pengeluaran anggaran |
| jenis | Jenis penerimaan dan pengeluaran anggaran |
| nilai\_anggaran | Nilai anggaran dalam juta Rp. |
| nilai\_realisasi | Nilai realisasi dalam juta Rp. |

1. Tabel Data Jumlah Pendapatan DKI Jakarta dari Pajak

Tabel ini berisi data jumlah pendapatan pajak DKI Jakarta tahun 2013. Berikut adalah penjelasan atribut pada tabel ini:

Tabel III‑17 Data Jumlah Pendapatan DKI Jakarta dari Pajak

| Atribut | Keterangan |
| --- | --- |
| nama\_kode\_rekening | Nama kode rekening |
| kategori | Kategori pendapatan/sumber pendapatan |
| tahun | Tahun |
| DPA | Nilai dari dokumen pelaksanaan anggaran (dalam Rp.) |

1. Tabel Data Anggaran dan Realisasi Penerimaan dan Pengeluaran

Tabel ini berisi anggaran dan realisasi penerimaan dan pengeluaran Pemerintah Provinsi DKI Jakarta tahun 2003-2012. Berikut adalah penjelasan atribut pada tabel ini:

Tabel III‑18 Data Anggaran dan Realisasi Penerimaan dan Pengeluaran

| Atribut | Keterangan |
| --- | --- |
| tahun | Tahun |
| anggaran\_realisasi | Anggaran yang direalisasikan |
| pos\_anggaran | Posisi anggaran |
| penjelasan\_pos\_anggaran | Penjelasan posisi anggaran |
| nilai | Nilai dalam juta Rp. |

1. Tabel Data Realisasi APBD DKI Jakarta per SKPD

Tabel ini berisi mengenai realisasi APBD DKI Jakarta per SKPD November 2013. Berikut adalah penjelasan atribut pada tabel ini:

Tabel III‑19 Data Realisasi APBD DKI Jakarta per SKPD

| Atribut | Keterangan |
| --- | --- |
| komisi | Komisi yang berkaitan |
| bidang | Bidang |
| lokasi\_id | ID Lokasi |
| id\_skpd | ID dari SKPD |
| kode\_skpd | Kode dari SKPD |
| nama\_skpd | Nama dari SKPD |
| nilai\_anggaran | Nilai Anggaran |
| nilai\_sp2d | Nilai dari surat perintah pencairan dana (SP2D) |
| persen\_sp2d | Persentase nilai SP2D terhadap nilai anggaran |
| nilai\_spj | Nilai dari dari Surat Pertanggung Jawaban (SPJ) |
| persen\_spj | Persen nilai SPJ terhadap nilai anggaran |
| persen\_rencana | Persentase dari rencana |
| persen\_pelaksanaan | Persentase dari yang terlaksana |

1. Tabel Data APBD DKI Jakarta Berdasarkan Kegiatan

Tabel ini berisi menganai APBD DKI Jakarta berdasarkan kegiatan tahun 2014. Berikut adalah penjelasan atribut pada tabel ini:

Tabel III‑20 Data APBD DKI Jakarta Berdasarkan Kegiatan

| Atribut | Keterangan |
| --- | --- |
| unit | Nomor unit |
| skpd\_nama | Nama SKPD |
| urusan | Kode urusan |
| nama\_urusan | Nama urusan |
| program | Nomor program |
| nama\_program | Nama program |
| no\_kegiatan | Nomor kegiatan |
| nama\_kegiatan | Nama kegiatan |
| nilai | Nilai anggaran |
| kegiatan\_id | ID kegiatan |
| skpd\_kode | Kode SKPD tahun 2013 |
| program\_kode | Kode program |
| realisasi | Realisasi anggaran |
| persen\_realisasi | Realisasi anggaran dalam persentase |
| fisik | Tag fisik |

1. Tabel Data Realisasi APBD DKI Jakarta Per Kegiatan

Tabel ini berisi mengenai realisasi APBD per kegiatan November 2013. Berikut adalah penjelasan atribut pada tabel ini:

Tabel III‑21 Data Realisasi APBD DKI Jakarta per Kegiatan

| Atribut | Keterangan |
| --- | --- |
| kategori | Nomor Kategori |
| id\_lokasi | ID Lokasi |
| komisi | Komisi yang berkaitan |
| id\_skpd | ID dari SKPD |
| kode\_skpd | Kode dari SKPD |
| nama\_skpd | Nama SKPD |
| kode\_urusan | Kode urusan |
| nama\_urusan | Nama urusan |
| no\_program | Nomor program |
| kode\_program | Kode program |
| nama\_program | Nama program |
| teks\_pegu | Penjelasan tambahan dari program |
| keyword1 | Kata kunci yang berkaitan dengan program 1 |
| keyword2 | Kata kunci yang berkaitan dengan program 2 |
| keyword3 | Kata kunci yang berkaitan dengan program 3 |
| id\_kegiatan | ID kegiatan |
| no\_kegiatan | Nomor kegiatan |
| nama\_kegiatan | Nama kegiatan |
| nilai\_anggaran | nilai anggaran yang berkaitan |
| volume | Volume |
| satuan | Satuan dari volume |
| nilai\_sp2d | Nilai dari surat perintah pencairan dana (SP2D) |
| persen\_sp2d | Persentase nilai SP2D terhadap total anggaran |
| nilai\_spj | Nilai dari surat pertangung jawaban |
| persen\_spj | Persentase nilai SPJ terhadap total anggaran |
| persen\_rencana | Persentase dari rencana |
| persen\_pelaksanaan | Persentase dari yang terlaksana |

1. Tabel Data Ringkasan Anggaran dan Realisasi, Pendapatan, Belanja, dan Pembiayaan DKI Jakarta

Tabel ini berisi ringkasan anggaran dan realisasi, pendapatan, belanja, dan pembiayaan di Provinsi DKI Jakarta pada tahun 2011 dan 2012. Berikut adalah penjelasan atribut pada tabel ini:

Tabel III‑22 Data Ringkasan Anggaran dan Realisasi,  
Pendapatan, Belanja, dan Pembiayaan

| Atribut | Keterangan |
| --- | --- |
| tahun | Tahun |
| akun | Keterangan anggaran |
| jenis | Jenis penerimaan dan pengeluaran anggaran |
| anggaran | Nilai anggaran dalam juta Rp. |
| realisasi | Nilai realisasi dalam juta Rp. |

### Analisis Segmentasi Pengguna Informasi

Analisis segmentasi pengguna informasi dilakukan dengan cara menentukan *early adoptor user*. Penentuan *early adoptor user* dilakukan dengan cara melakukan wawancara kepada seorang ahli dibidang ekonomi dan keuangan. Kegiatan wawancara dilakukan dengan Ibu Sri Dewi Anggadini, SE., M.Si, dosen tetap Fakultas Ekonomi di Universitas Komputer Indonesia (UNIKOM). Menurut beliau, melihat dari dataset yang disajikan dalam penelitian ini, segmentasi pengguna informasi secara garis besar dibedakan menjadi dua kelompok. Kelompok pertama adalah kelompok ekonomi makro dan kekompok kedua dalah kelompok ekonomi mikro. Baik pelaku ekonomi makro maupun pelaku ekonomi mikro, memiliki pandangan dan kebutuhan informasi yang independen. Pelaku ekonomi makro adalah mereka yang melihat ekonomi secara lebih luas, sedangkan pelaku ekonomi mikro adalah mereka yang melihat ekonomi sebagai permintaan dan penawaran terhadap barang dan jasa. Pelaku ekonomi makro secara garis besar memerlukan informasi yang memengaruhi banyak masyarakat atau yang lebih abstrak, seperti perusahaan atau pasar. Sedangkan pelaku ekonomi mikro secara garis besar memerlukan informasi yang lebih detail, seperti bagaimana penentuan satu harga akan menentukan penawaran dan permintaan barang atau jasa selanjutnya.

Tujuan segmentasi pengguna adalah untuk membedakan informasi yang akan dikelola ditinjau dari data ekonomi dan keuangan daerah yang diambil dari tempat penyimpanan data Pemprov DKI Jakarta. Setelah melakukan segmentasi pengguna untuk visualisasi data, yaitu kelompok yaitu pelaku ekonomi makro dan pelaku ekonomi mikro. Selanjutnya segemntasi pengguna akan dimanfaatkan untuk memaksimalkan visualisasi data berdasarkan sudut pandang setiap segmen.

### Analisis Kebutuhan Informasi Berdasarkan Segmentasi Pengguna

Analisis kebutuhan informasi pengguna berdasarkan segmentasi pengguna dilakukan untuk mencari pertanyaan-pertanyaan atau masalah yang terdapat pada setiap segmen. Selanjutnya pertanyaan tersebut akan dicocokan berdasarkan dataset ekonomi dan keuangan yang digunakan pada penelitian ini. Pertanyaan yang sesuai dengan dataset penelitian akan digunakan dalam proses visualisasi data.

#### Informasi Strategis Pelaku Ekonomi Makro

Dalam ekonomi makro, variabel-variabel ekonomi dilihat secara keseluruhan. Variabel tersebut seperti pendapatan nasional, kesempatan kerja atau pengangguran, jumlah uang beredar, laju inflasi, pertumbuhan ekonomi, dan neraca pembayaran internasional [21]. Dalam ekonomi makro, terdapat lima pelaku utama yang menjalankan produk ekonomi [22]. Kelima pelaku tersebut yaitu:

1. *Households*
2. *Business*
3. *Government*
4. *Foreign Countries*
5. *Financial*

Permasalahan utama dalam ekonomi makro mempelajari masalah-masalah sebagai berikut:

1. Sejauh mana sumber daya telah dimanfaatkan di dalam kegiatan ekonomi.
2. Sejauh mana perekonomian dalam keadaan stabil khususnya dibidang moneter.
3. Sejauh mana perekonomian mengalami pertumbuhan, dan pengaruhnya (distribusi) ke pelaku ekonomi.

Adapun dataset yang berkaitan dengan ekonomi makro yang terdapat di tempat penyimpanan *open data* Pemerintah Provinsi DKI Jakarta adalah sebagai berikut:

1. **Data Ketenaga Kerjaan**
   1. Tabel III‑6 Data Ikhtisar Statistik Antar Kerja DKI Jakarta
   2. Tabel III‑7 Data Upah Minimum Provinsi dan Inflasi
2. **Data Ekspor-Impor**
   1. Tabel III‑4 Data Ekspor Impor DKI Jakarta
3. **Data Pertumbuhan Ekonomi**
   1. Tabel III‑1 Data Laju Pertumbuhan Ekonomi Jakarta dan Nasional
   2. Tabel III‑2 Data Tingkat Inflasi Jakarta dan Nasional
   3. Tabel III‑3 Data Komponen Inflasi Jakarta
   4. Tabel III‑5 Data Struktur Ekonomi Jakarta Berdasarkan Sektor
   5. Tabel III‑13 Data Pendapatan Perkapitra Jakarta dan Nasional

#### Informasi Strategis Pelaku Ekonomi Mikro

Mikro ekonomi membahas bagaimana pengambilan keputusan seorang individu dengan sumber daya individu tersebut yang terbatas dalam memenuhi kebutuhannya. Variabel ekonomi mikro berhubungan dengan variabel-variabel permintaan dan penawaran, yaitu variabel harga dan variabel penjualan. Variabel tersebut ditentukan oleh seberapa besar permintaan asar [23]. Dalam ekonomi mikro, pelaku yang terlibat yaitu [22]:

1. *Households*
2. *Business Firms*

Permasalahan utama dalam ekonomi mikro mempelajari masalah-masalah sebagai berikut:

1. Bagaimana harga suatu komoditas.
2. Bagaimana permintaan dan penawaran suatu komoditas, perilaku konsumen, perilaku produsen, pasar, dan untung/rugi.
3. Bagaimana cara mengalokasikan sumber daya agar dapat mencapai kombinasi yang tepat.

Adapun dataset yang berkaitan dengan ekonomi mikro yang terdapat di tempat penyimpanan *open data* Pemerintah Provinsi DKI Jakarta adalah sebagai berikut:

1. **Data Anggaran Pendapatan dan Belanja Daerah**
   1. Tabel III‑11 Data APBD Perubahan DKI Jakarta 2014
   2. Tabel III‑12 Data APBD Penetapan DKI Jakarta 2014
   3. Tabel III‑14 Data Anggaran Pendapatan Belanja Daerah
   4. Tabel III‑15 Data Anggaran Belanja Langsug dan Tidak Langsung
   5. Tabel III‑16 Data Anggaran dan Realisasi Pendapatan  
      Daerah menurut Sumber dan Jenis Pengeluaran
   6. Tabel III‑17 Data Jumlah Pendapatan DKI Jakarta dari Pajak
   7. Tabel III‑18 Data Anggaran dan Realisasi Penerimaan dan Pengeluaran
   8. Tabel III‑19 Data Realisasi APBD DKI Jakarta per SKPD

### Analisis Algoritma

Analisis algoritma adalah bagian dimana untuk setiap kelompok dataset dipetakan terhadap algoritma tertentu. Terdapat kelompok dataset ketenaga kerjaan, dataset pertumbuhan ekonomi, dan dataset anggaran pendapatan belanja daerah.

#### Analisis Algoritma untuk Data Ketenaga Kerjaan

#### Analisis Algoritma untuk Data Pertumbuhan Ekonomi

Pada kelompok data pertumbuhan ekonomi, terdapat lima dataset yang digunakan, tetapi dataset yang akan diolah dengan algoritma sebelum dilakukan visualisasi hanya tiga dataset, yaitu data pertumbuhan ekonomi Jakarta dan nasional, data pendapatan perkapitra Jakarta dan nasional, dan data inflasi Jakarta dan nasional. Dua dataset lainnya yaitu data komponen inflasi Jakarta dan struktur ekonomi Jakarta berdasarkan sektor akan langsung diolah menjadi bentuk visual.

Di bawah ini adalah contoh data yang terdapat pada dataset pertumbuhan ekonomi Jakarta dan nasional, pendapatan perkapita Jakarta dan nasional, dan data inflasi Jakarta dan nasional.

Tabel III‑23 Data Pertumbuhan Ekonomi Jakarta dan Nasional

| tahun | persen\_tumbuh\_jakarta (%) | persen\_tumbuh\_nasional (%) |
| --- | --- | --- |
| 2006 | 5.95 | 5.5 |
| 2007 | 6.44 | 6.35 |
| 2008 | 6.23 | 6.01 |
| 2009 | 5.02 | 4.63 |
| 2010 | 6.5 | 6.2 |
| 2011 | 6.73 | 6.48 |
| 2012 | 6.53 | 6.23 |

Tabel III‑24 Data Pendapatan Perkapita Jakarta dan Nasional

| tahun | perkapita\_jakarta (Juta Rp.) | perkapita\_nasional (Juta Rp.) |
| --- | --- | --- |
| 2006 | 55.98 | 15.33 |
| 2007 | 62.49 | 17.5 |
| 2008 | 74.16 | 21.7 |
| 2009 | 82.15 | 23.9 |
| 2010 | 89.72 | 27.1 |
| 2011 | 100.98 | 30.4 |
| 2012 | 110.46 | 33.3 |

Tabel III‑25 Data Inflasi Jakarta dan Nasional

| tahun | inflasi\_jakarta (%) | inflasi\_nasional (%) |
| --- | --- | --- |
| 2006 | 6.03 | 6.6 |
| 2007 | 6.04 | 6.59 |
| 2008 | 11.11 | 11.06 |
| 2009 | 2.34 | 2.78 |
| 2010 | 6.21 | 6.96 |
| 2011 | 3.97 | 3.79 |
| 2012 | 4.52 | 4.3 |

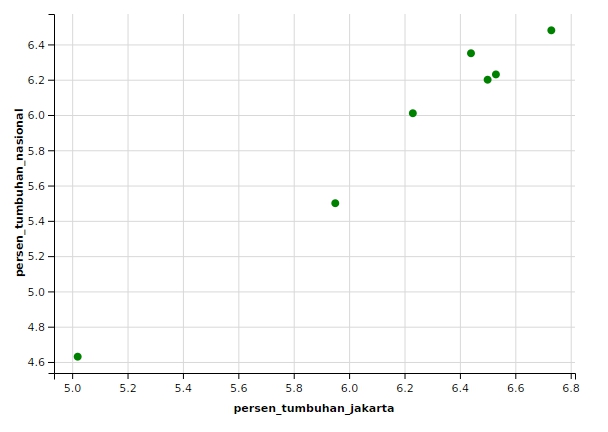
Berdasarkan analisis kebutuhan informasi pada bagian sebelumnya, Tabel III‑23, Tabel III‑24, dan Tabel III‑25 termasuk kedalam segmen ekonomi makro. Salah satu masalah yang dibahas dalam ekonomi makro yaitu:

* Sejauh mana perekonomian mengalami pertumbuhan, dan pengaruhnya (distribusi) ke pelaku ekonomi.

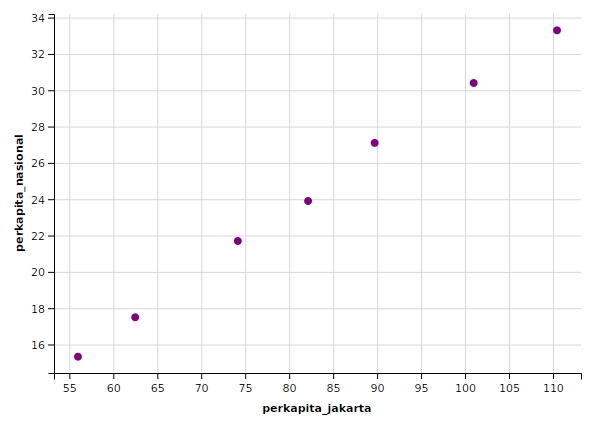
Mengacu pada masalah tersebut, penting dibutuhkan peramalan terhadap data yang ada. Peramalan ini akan memberikan model peramalan tentang pergerakan data yang ada.

##### Peramalan dengan Regresi Linear

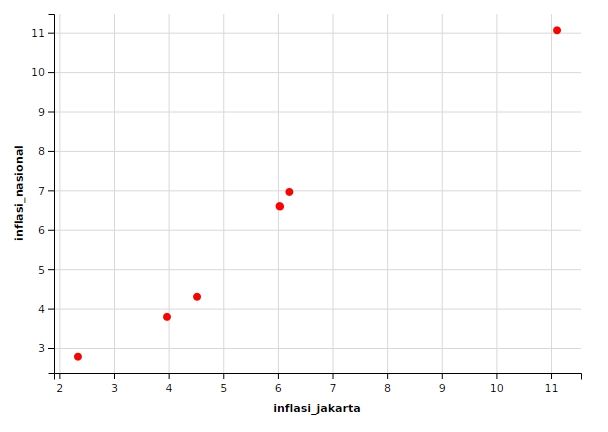
Memprediksi *y* (*depenent variable*) berdasarkan acuan *x* (*indevedent variable*) dapat dilakukan dengan membuat garis yang sangat dekat dengan setiap kordinat pada titik (*x, y*). Prediksi *dependent variable* dapat dilakukan dengan regresi linier. Regresi adalah suatu algoritma yang dapat memprediksi keterkaitan antar *variable*. Berikut adalah *scatter plot* untuk Tabel III‑23, Tabel III‑24, dan Tabel III‑25.



Gambar III‑1 *Scatter Plot* untuk Tabel III‑23



Gambar III‑2 *Scatter Plot* untuk Tabel III‑24



Gambar III‑3 *Scatter Plot* untuk Tabel III‑25

Dalam teori korelasi, *r* (*correlation coefficient*) adalah koefisien yang menghitung seberapa kuat kerapatan titik-titik pada *scatter plot*, rentang *r* adalah [-1 – 1]. Koefisien yang mengarah ke -1 atau -1 menunjukan titik-titik pada *scatter plot* sangat dekat dan cenderung miring negatif, sebaliknya 1 atau 1 menunjukan titik-titik pada *scatter plot* sangat dekat dan cenderung miring positif. *r* dapat dihitung dengan formula sebagai berikut:

(1)

Sehingga didapatkan nilai koefisien *r untuk* Tabel III‑23 sebesar 0.988587461~ atau 0.989, Tabel III‑24 sebesar 0.9992548066~ atau 0.999 dan, Tabel III‑25 sebesar 0.9892875154~ atau 0.989.

Tahap selanjutnya membuat garis regresi adalah menentukan koefisien kemiringan atau gradient garis (*b*), dan satu kordinat acuan untuk melakukan *ploting* garis. Kemiringan garis regresi dapat dihitung dengan mengalikan koefisien korelasi (*r*) dengan rasio dengan

(2)

Untuk Tabel III‑23 nilai kemiringan dari garis regresi adalah 1.1118803933~ atau 1.112; Tabel III‑24 memiliki tingkat kemiringan 0.3322060892~ atau 0.332; dan Tabel III‑25 memiliki tingkat kemiringan 0.988872594~ atau 0.989.

Kordinat atau titik yang paling tepat menjadi acuan garis regresi adalah titik yang paling dekat relatif dengan titik-titik lain, titik tersebut adalah rata-rata dari setiap titik, (). Sehingga didapatkan persamaan garis regresi sebagai berikut

(3)

*a* adalah *intercept* yang menunjukan seberapa jauh garis regresi dari kordinat (0,0). Dengan persamaan aljabar sederhana, nilai *intercept a* dapat dihitung dengan .

Segingga didapatkan model garis regresi untuk variabel *x* terhadap *y* untuk Tabel III‑23, ketiga model tersebut yaitu:

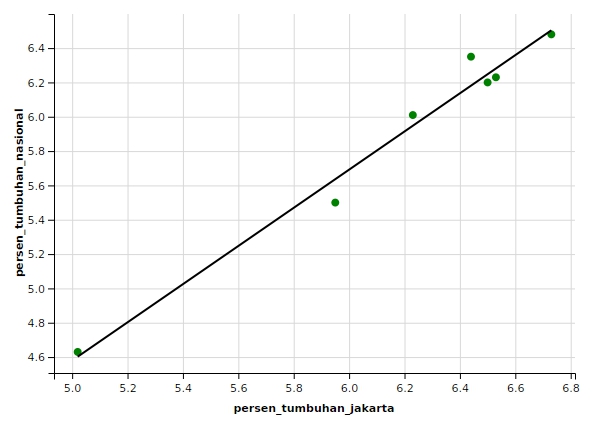
(4)

(5)

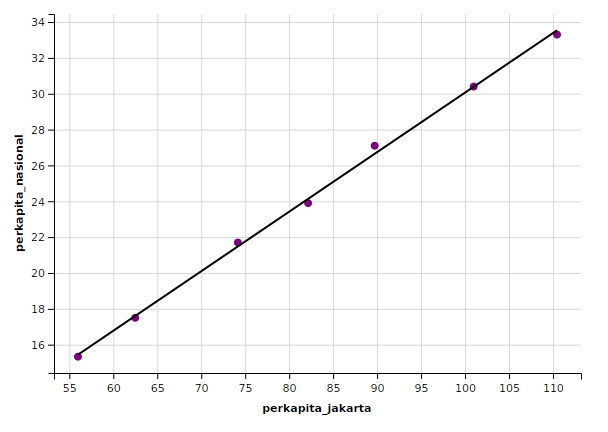
(6)

Model di atas digunakan untuk memprediksi variabel *y* terhadap variabel *x*.

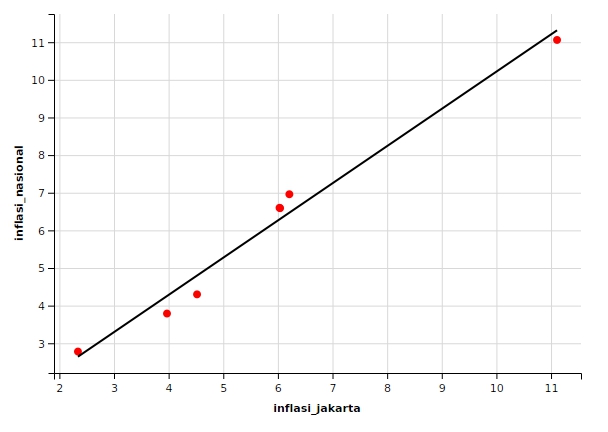
Untuk mengukur akurasi garis regresi dengan total variansi titik *x*, maka dihitung nilai dari *coefficient of determination (r2).* Nilai dari *r2* untuk Tabel III‑23 adalah 0.977305168~ atau 0.997; Tabel III‑24 adalah 0.9985101686~ atau 0.999; dan Tabel III‑25 adalah 0.978689788~ atau 0.979. Artinya persamaan (4) menunjukan 99.7 % total variasi titik *y* akurat dengan total variasi titik *x*; persamaan (5) menunjukan 99.9% total variasi titik *y* akurat dengan total variasi titik *x*; dan persamaan (6) menunjukan 97.9% total variasi titik *y* akurat dengan total variasi titik *x*. Berikut adalah *scatter plot* dengan model garis linear untuk Tabel III‑23.

****

Gambar III‑4 *Scatter Plot* dengan Modeal Linear Tabel III 23



Gambar III‑5 *Scatter Plot* dengan Modeal Linear Tabel III‑24



Gambar III‑6 *Scatter Plot* dengan Modeal Linear Tabel III‑25

Selanjutnya adalah menghitung interval error dengan kondisi sebenarnya. Rho *(ρ)* adalah koefisien yang menunjukan sebarapa akurat prediksi dengan populasi sebenarnya. Titik estimasi *ρ* adalah *r.* *Confidence Interval ρ* untuk batas atas dihitung dengan menjumlahkan *r* dengan perkalian antara batas atas *standard error* *r* dengan *tα/2*. Sebaliknya, batas bawah dihitung dengan mengurangi *r* dengan perkalian antara batas atas *standard error* *r* dengan *tα/2*. *α* level yang digunakan adalah 0.05 *two tail*. Formula untuk menghitung *confidence interval ρ* adalah sebagai berikut:

(7)

(8)

Sehingga didapatkan *confident interval* untuk .05 Tabel III‑23 adalah (0.9844664578~, 0.9927084642~) atau (0.984, 0.993). Nilai *r* adalah 0.989, *r* jatuh diantara *confident interval* . Artinya persamaan (4) memprediksi 95% terjadi bukan karena acak (kebetulan). Untuk Tabel III‑24 , *confident interval* untuk adalah (0.9922459924~,1.0062636209~) atau (0.992, 1.00). Nilai *r* adalah 0.999, *r* jatuh diantara *confident interval*  *.* Artinya persamaan (5) memprediksi 95% terjadi bukan karena acak (kebetulan). Untuk tabel Tabel III‑25, *confident interval* untuk adalah (0.9627799483~,1.0157950824) atau (0.962,1.016). Nilai *r* adalah 0.989, *r* jatuh diantara *confident interval*  *.* Artinya persamaan (6) memprediksi 95% terjadi bukan karena acak (kebetulan).

### *Data Preprocessing*

Pada bagian *data preprocessing* data yang terdapat pada *flat file* akan dilakukan pengolahan dan pembersihan data terlebih dahulu sebelum disimpan kedalam basis data. Adapun langkah-langkah *data preprocessing* adalah mengikuti *data cleaning* menurut Megan Squire [24] dan Arthur D. Chapman [25]. Langkah-langkah *data preprocessing* adalah sebagai berikut:

1. Pemeriksaan data dan pendeteksian kesalahan

Tahap ini adalah tahap dimana data yang tidak konsisten dicari dan diperbaiki. Dataset yang digunakan untuk tahap ini adalah dataset pada kelompok Data Anggaran Pembelanjaan Daerah.

1. Validasi data

Tahap ini adalah tahap dimana data pada setiap atribut tabel ditentukan tipe datanya, seperti tipe data untuk tanggal, waktu, karakter/string, dan nomor.

1. Koreksi kesalahan

Koreksi kesalahan adalah bagian dimana data pada setiap atribut diperiksa kembali apakah sudah sesuai dengan data yang dimaksud. Contoh: konversi tipe data — sperti string ke nomor untuk pengolahan matematis, atau sebaliknya; membedakan antara null dan 0; atau penentuan *character encoding*.

### Pemetaan Visualisasi Data

Membuat visualisasi membutuhkan beberapa tahap penilaian yang yang bisa dibedakan. Thap-tahap yang tidak boleh dilewatkan seperti menentukan pertanyaan yang hendak dicari jawabannya, identifikasi data yang diperlukan, dan memilih *visual encoding* untuk dipetakan ke nilai-nilai grafis, seperti posisi, ukuran, bentuk, dan warna [26]. Tantangan dari pemetaan visualisasi data adalah menentukan visualisasi yang paling tepat dari dataset yang diberikan. Untuk menjawab tantangan tersebut, pembuat visualisasi harus memahami betul bagaimana *encoding* yang berbeda seperti angka, kategori, dan *networks* dipetakan ke bentuk visualisasi seperti *bar chart, line chart, scatter plot,* dan posisi [26]. Berikut adalah pemetaan visualisasi data kedalam setiap dataset yang digunakan pada penelitian tugas akhir ini.

#### *Time-Series* Data

*Time-series* data adalah data yang nilainya berubah setiap periode waktu. Kebutuhan visualisasi pada data ini biasanya membandingkan satu data pada satu periode waktu dengan periode waktu lainnya. Untuk *time-series* data, jenis grafik visual yang dipakai adalah *index charts, horizon graph,* atau temporal. Dataset yang menggunakan visualisasi *time-series* data diantaranya yaitu:

1. Tabel III‑1 Data Laju Pertumbuhan Ekonomi Jakarta dan Nasional
2. Tabel III‑2 Data Tingkat Inflasi Jakarta dan Nasional
3. Tabel III‑3 Data Komponen Inflasi Jakarta
4. Tabel III‑4 Data Ekspor Impor DKI Jakarta
5. Tabel III‑7 Data Upah Minimum Provinsi dan Inflasi
6. Tabel III‑25 Data Inflasi Jakarta dan Nasional

#### Distribusi Statistik

Visualisasi distribusi statistik didesain untuk membantu analis dalam memahami properti statistik. Analis membutuhkan visualisasi yang sesuai dengan model untuk melakukan prediksi atau hipotesis [26]. Grafik visual yang digunakan untuk distribusi statistik yaitu *scatter plot,* *maps,* atau *cartograms.* Dataset yang menggunakan visualisasi distribusi statistik yaitu:

1. Tabel III‑1 Data Laju Pertumbuhan Ekonomi Jakarta dan Nasional
2. Tabel III‑2 Data Tingkat Inflasi Jakarta dan Nasional
3. Tabel III‑4 Data Ekspor Impor DKI Jakarta
4. Tabel III‑7 Data Upah Minimum Provinsi dan Inflasi
5. Tabel III‑24 Data Pendapatan Perkapita Jakarta dan Nasional

#### Hirarki

*Flat file* yang memiliki data dalam bentunk angka (bisa dihitung), hampir semua jenis data seperti ini bisa divisualisasikan ke dalam hirarki [26]. Grafik visual yang termasuk hirarki yaitu *tree,* diagram node,atau *treemap*. Dataset yang menggun hirarki yaitu:

1. Tabel III‑8 Data Laporan Realisasi Anggaran
2. Tabel III‑10 Data Anggraran Belanja DKI Jakarta Per Kegiatan
3. Tabel III‑11 Data APBD Perubahan DKI Jakarta 2014
4. Tabel III‑12 Data APBD Penetapan DKI Jakarta 2014
5. Tabel III‑14 Data Anggaran Pendapatan Belanja Daerah
6. Tabel III‑17 Data Jumlah Pendapatan DKI Jakarta dari Pajak
7. Tabel III‑20 Data APBD DKI Jakarta Berdasarkan Kegiatan
8. Tabel III‑21 Data Realisasi APBD DKI Jakarta per Kegiatan
9. Tabel III‑22 Data Ringkasan Anggaran dan Realisasi, Pendapatan, Belanja, dan Pembiayaan

## Analisis Spesifikasi Kebutuhan Perangkat Lunak

Bagian analisis spesifikasi kebutuhan perangkat lunak menentukan kebutuhan-kebutuhan apa saja yang diperlukan untuk membangkun perangkat lunak visualisasi interaktif. Analisis ini dibagi menjadi dua bagian, tahap analisis kebutuhan non-fungsionaldan tahap analisi kebutuhan fungsional.

### Analisis Kebutuhan Non-Fungsional

Tujuan analisis kebutuhan non-fungsional adalah untuk menentukan spesifikasi kebutuhan sistem yang akan dibangun. Terdapat dua bagian dalam analisis non-fungsional, analisis perangkat keras dan analisis perangkat lunak. Analisis Kebutuhan Perangkat Keras

Analisis kebutuhan perangkat keras adalah tahap analisis terhadap perangkat keras yang akan digunakan dalam menjalankan pernagkat lunak yang dibangun. Spesifikasi perangkat keras yang digunakan pada perangkat lunak visualisasi data interaktif ini adalah sebagai berikut:

1. *Processor: 32/64 bit processor*
2. *Memory: >= 512 MB*
3. *Disk: :> 5 GB*
4. *Network: 10 MBps Uplink/Downlink*

Analisis perangkat keras mengikuti spesifikasi minimum dari *dependency* perangkat lunak yang digunakan untuk menjalankan perangkat lunak visualisasi data interaktif.

#### Analisis Kebutuhan Perangkat Lunak

Analisis kebutuhan perangkat lunak adalah tahap analisis terhadap perangkat lunak yang digunakan untuk menjalankan perangkat lunak visualisasi data interaktif. *Dependency* perangkat lunak untuk menjalankan perangkat lunak visualisasi data interaktif adalah sebagai berikut:

1. Aplikasi dibangun di atas sistem operasi Linux, distro Ubuntu 14.04.x server.
2. Aplikasi menggunakan server node.js 4.x.x
3. Aplikasi dijalankan menggunakan basis data NoSQL MongoDB™ 3.0.x

### Analisis Kebutuhan Fungsional

#### Diagram Konteks

#### Data Flow Diagram

##### DFD Level 1

##### DFD Level 2 ..

##### DFD Level 2..

dst

#### Spesifikasi Proses

#### Kamus Data DFD

#### Skema Basis Data

#### Struktur Tabel

## Perancangan Arsitektur

### Perancangan Struktur Menu

### Perancangan Antarmuka

#### Perancangan Tampilan Program

#### Perancangan Tampilan Pesan

#### Jaringan Semantik

# IMPLEMENTASI DAN PENGUJIAN

Implementasi Sistem

Implementasi Perangkat Keras

Implementasi Perangkat Lunak

Implementasi Form

Implementasi Basis Data

Pengujian Perangkat Lunak

Rencana Pengujian

Pengujian Black Box

Pengujian Beta

Pengujian Sample

# 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. Silberschatz, H. F. Korth, and S. Sudarshan, *Database System Concepts - 6th. ed.*, vol. 4. 2011.

[10] G. Vaish, *Getting started with NoSQL*. 2013.

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

[12] “Visual Encoding | Targetprocess - Visual management software.” [Online]. Available: https://www.targetprocess.com/articles/visual-encoding/. [Accessed: 17-Dec-2015].

[13] 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.

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

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

[16] “Udacity - Intro to Inferential Statistics - Statistical Significance.” [Online]. Available: https://www.udacity.com/course/viewer#!/c-ud201/l-1330628575/m-230229089. [Accessed: 16-Dec-2015].

[17] “R-Squared: Sometimes, a Square is just a Square.” [Online]. Available: http://blog.minitab.com/blog/statistics-and-quality-data-analysis/r-squared-sometimes-a-square-is-just-a-square. [Accessed: 16-Dec-2015].

[18] “MATH 241: Materials - confidence interval for r.” [Online]. Available: http://turner.faculty.swau.edu/mathematics/math241/materials/rconf/. [Accessed: 16-Dec-2015].

[19] R. M. Forte, *Mastering Predictive Analytics with R*. 2015.

[20] C. O’Neil and R. Schutt, *Doing Data Science: Straight Talk from the Frontline*. 2013.

[21] “Microeconomics and Macroeconomics | Microeconomics.” [Online]. Available: https://courses.candelalearning.com/microecon/chapter/microeconomics-and-macroeconomics/#m48592-fs-idp49079232. [Accessed: 01-Dec-2015].

[22] Boundless, “Macroeconomics,” 21-Jul-2015. [Online]. Available: https://www.boundless.com/economics/textbooks/boundless-economics-textbook/principles-of-economics-1/differences-between-macroeconomics-and-microeconomics-44/macroeconomics-162-12260/. [Accessed: 01-Dec-2015].

[23] Boundless, “Microeconomics,” 21-Jul-2015. [Online]. Available: https://www.boundless.com/economics/textbooks/boundless-economics-textbook/principles-of-economics-1/differences-between-macroeconomics-and-microeconomics-44/microeconomics-163-12261/. [Accessed: 01-Dec-2015].

[24] M. Squire, *Clean Data*. Birmingham: Packet Publishing Ltd., 2015.

[25] A. D. Chapman, “Principles and\nmethods of data cleaning - primary\nspecies and species-occurrence data,” *Rep. Glob. Biodivers. Inf. Facil.*, no. version 1.0, p. 72, 2005.

[26] J. Heer, M. Bostock, and V. Ogievetsky, “VIISUALIZATION A Tour through the Visualization Zoo A survey of powerful visualization techniques , from the obvious to the obscure,” *Commun. ACM*, vol. 53, no. 5, pp. 59–67, 2010.