

VISUALISASI DATA INTERAKTIF DATA TERBUKA PEMERINTAH PROVINSI DKI JAKARTA: TOPIK EKONOMI DAN KEUANGAN DAERAH

Nadiar Ahmad Syaripul dan Adam Mukharil Bachtiar

Teknik Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Komputer Indonesia,
Jalan Dipatiukur No. 112-116, Bandung, 40132, Indonesia

Email: nadiar429@email.unikom.ac.id, adam@email.unikom.ac.id

Abstract

Based on statistics from data.id, in the first quarter of 2016, there are 1,137 datasets distributed at 32 institutions and 18 groups in Indonesia. DKI Jakarta Province contributes to these data at the most, i.e. 714 datasets. A lot of accessible open datasets have an impact on the availability of valuable information that can be extracted to good use, for businesses, governments, and personal lives. To get the desired information, an exploratory data analysis is needed to make data more alive. The goal of this research is to provide a proper visualization of the given data. Data visualization is a way (perhaps a solution) to communicate abstract data, to aid in data understanding by leveraging human visual system. The result of this visualization is effective and engaging charts appropriate to the given data and can be run on mobile platforms.

Keywords: *visualization, open data, information, exploratory data analysis, mobile platform*

Abstrak

Berdasarkan statistik data.id, di kuartal pertama 2016, data terbuka yang tersedia di Indonesia sebanyak 1.137 dataset yang tersebar di 32 instansi dan 18 grup. Provinsi DKI Jakarta menjadi provinsi yang paling banyak memberikan kontribusi dataset, yaitu sebanyak 714. Banyaknya dataset terbuka yang bebas diakses berdampak pada informasi berharga yang bisa didapatkan dan dimanfaatkan untuk keperluan bisnis, pemerintahan, ataupun pribadi. Untuk mendapatkan informasi tersebut, analisis eksplorasi data diperlukan untuk membuat data lebih hidup. Tujuan penelitian ini membuat visualisasi yang tepat dari data yang diberikan. Visualisasi data adalah cara (solusi) mengkomunikasikan data abstrak, membantu pemahaman data dengan memanfaatkan sistem visual manusia. Hasil visualisasi ini adalah grafik yang efektif dan menarik yang tepat untuk data dan berjalan di atas platform *mobile*.

Kata Kunci: *visualisasi, data terbuka, informasi, analisis eksplorasi data, mobile platform*

1. Pendahuluan

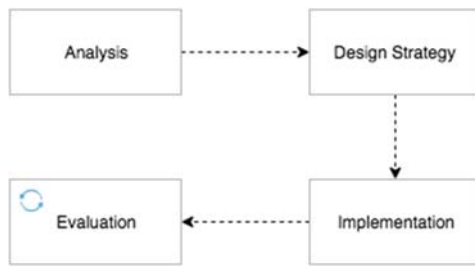
Pada era digital saat ini, menemukan data tidaklah begitu sulit; data (dataset) dapat ditemukan di internet. Dataset tersebut sangatlah besar, kompleks, dan sulit dianalisis dengan perangkat lunak *data processing* seperti *spreadsheet* [1]. Pemerintah Provinsi DKI Jakarta menyediakan dataset yang bebas didistribusikan, disimpan, dan dikelola. Dataset seperti ini dikenal dengan istilah "*open data*" (data terbuka) [2].

Data ekonomi dan keuangan daerah menggambarkan data tentang pertumbuhan ekonomi dan hasil kinerja pemerintah, data ini ada sebagai upaya transparansi pemerintah baik kepada pelaku ekonomi maupun pelaku non-ekonomi. Berdasarkan hasil kuesioner terbuka yang dilakukan pada tanggal 25 Agustus 2015 s.d. 27 Agustus 2015 dengan segmentasi responden pelaku UKM di Jakarta, 78.05 % responden tidak bisa mengambil *insight* dari da-

ta, 21.95% yang dapat mengambil *insight* dari data, dan 100% responden setuju jika visual analisis diperlukan. Kesimpulannya mayoritas responden setuju bahwa data yang disajikan dalam tempat penyimpanan data Pemerintah Provinsi DKI Jakarta masih sulit dibaca dan dimengerti.

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 [3].

Penelitian ini menggunakan pustaka *nvD3.js* dalam implementasi grafik dinamis dan interaktif. Grafik diimplementasikan di atas platform *hybrid* aplikasi *mobile* menggunakan *framework* *ionic 2*. Data yang digunakan dalam aplikasi *mobile* diam-



Gambar 1. Diagram Metode Penelitian

bil langsung dari *flat file* tempat penyimpanan data terbuka DKI Jakarta. Data tersebut diperbarui dengan rentang waktu tertentu menggunakan *cron-job*.

2. Metode

Tahapan Pengembangan Penelitian

Penelitian ini merupakan penelitian pengembangan dengan tahapan sebagaimana ditunjukkan pada Gambar 1.

Analysis

Langkah pertama yang dilakukan adalah melakukan analisis masalah, analisis kebutuhan pengguna, dan analisis pemilihan algoritma. Analisis masalah dilakukan untuk mencari tahu permasalahan pengguna *open data* topik ekonomi dan keuangan daerah. Analisis kebutuhan pengguna dilakukan untuk memastikan kebutuhan pengguna terhadap data. Sementara itu, analisis pemilihan algoritma dilakukan untuk memilih algoritma paling efisien sesuai sumber daya yang tersedia dan kebutuhan pengguna.

Design Strategy

Berdasarkan analisis yang diperoleh, kemudian dibuatlah strategi pemecahan masalah, strategi ini meliputi strategi perancangan data dan perancangan arsitektur.

Implementation

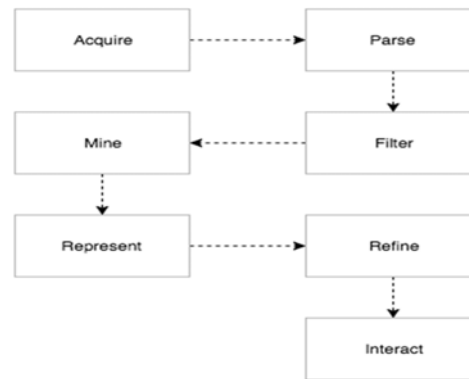
Design Strategy diimplementasikan pada data dan teknologi arsitektur yang sudah dipilih.

Evaluation

Evaluasi berkelanjutan yang dilakukan pengembangan untuk terus menyempurnakan sistem.

Batasan Masalah

Karena data yang sangat banyak dan terpisah-pisah dilakukan batasan dalam pengambilan data, yaitu sebanyak 12 data pada topik ekonomi dan keuangan



Gambar 2. Proses visualisasi data menurut Ben Fry dalam buku yang berjudul “Visualizing Data”

an daerah, dengan memilah data yang paling populer berdasarkan *activity stream* repositori. Adapun data yang digunakan tertera pada Lampiran Tabel I.

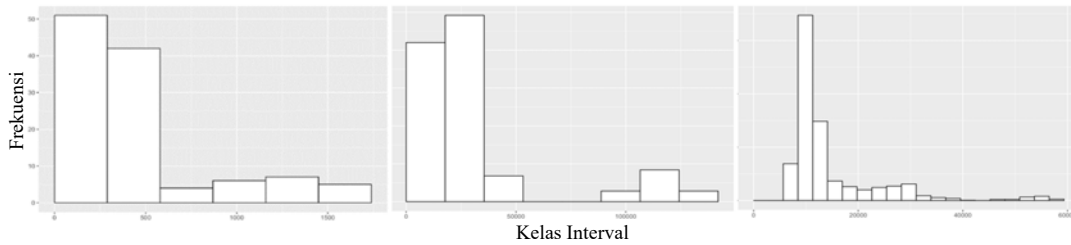
Proses Visualisasi Data

Visualisasi data sama halnya seperti berkomunikasi, berhasil atau tidaknya komunikasi ditentukan oleh bagaimana cara pembicara menyampaikan informasi yang diberikan kepada lawan komunikasi. Visualisasi yang baik tentu terfokus, memberikan jawaban yang jelas, dan tidak terlalu detail. Untuk mencapai visualisasi yang baik dilakukan proses visualisasi data [4], proses visualisasi data yang digunakan adalah sebagaimana ditunjukkan pada Gambar 2.

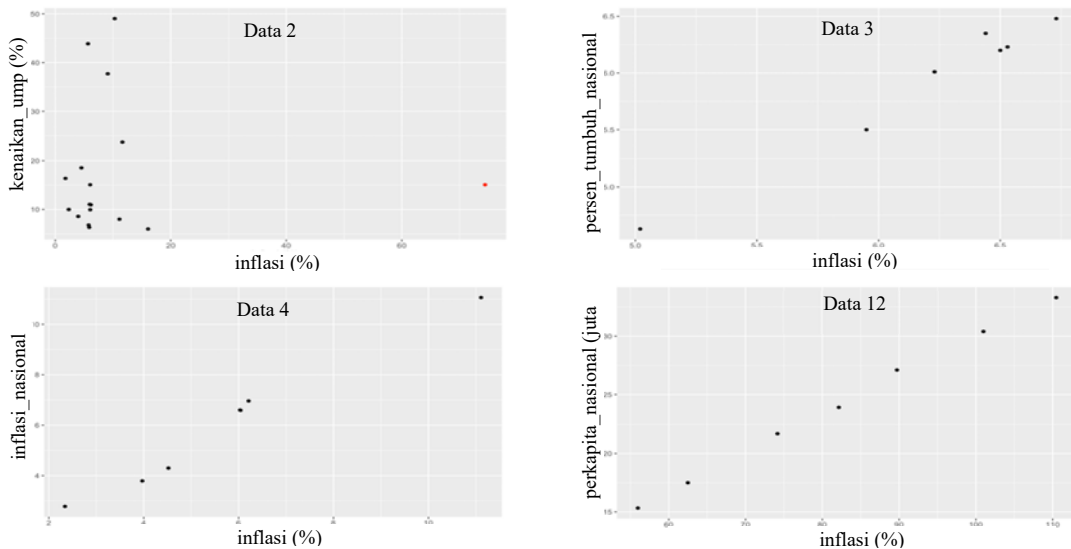
Acquire meliputi proses pengambilan data, baik dari internet ataupun dari *disk* lokal. *Parse* merupakan proses memberikan struktur data yang tepat pada data yang akan diolah. *Filter* merupakan proses memilah-milah data yang sesuai dengan tujuan yang akan dicapai. *Mine* meliputi proses melakukan analisis statistik atau *data-mining* terhadap data untuk mencapai tujuan. Proses ini bersifat pilihan. *Represent* merupakan proses merubah data yang berbentuk teks menjadi grafik dasar yang sesuai. *Refine* adalah proses memperkaya grafik dasar dengan berbagai *visual encoding*. Proses ini bersifat pilihan. *Interact* adalah proses merubah grafik statis menjadi grafik dinamis. Proses ini juga bersifat pilihan.

Exploratory Data Analysis

Exploratory Data Analysis (EDA) adalah proses menggali intuisi (memahami) terhadap data yang digunakan, apakah data ini dan untuk apa data ini [5]. Pada penelitian ini, EDA digunakan untuk menjawab kebutuhan informasi dari setiap data (*ex-*



Gambar 3. Histogram untuk Data 8, Data 10, dan Data 11



Gambar 4. Scatter plot untuk Data 2, 3, 4, dan 12

tract information from the data). Kebutuhan informasi tersebut selanjutnya dipetakan terhadap metode statistik atau algoritma tertentu (pemecahan masalah) sebelum dilakukan visualisasi. Adapun kebutuhan informasi dan metode pemecahan masalah ditunjukkan pada Lampiran Tabel II.

Untuk data yang tidak terdapat pada Lampiran Tabel II, visualisasi langsung didapatkan tanpa melalui proses pemecahan masalah (proses *Mine*).

3. Hasil dan Analisis

Data 1 dan Data 9 menunjukkan angka-angka statistik secara eksplisit, sehingga permasalahan dapat dijawab secara langsung dan sederhana melalui statistik dalam bentuk *bar chart* atau *pie chart* [6].

Histogram

Histogram digunakan pada Data 8, Data 10, dan Data 11 (Gambar 3). Histogram dipakai untuk menggambarkan distribusi terhadap suatu *field* pada data tersebut [7]. Panjang interval pada histogram dihitung dengan formula *scott* [8], yaitu:

$$h = \frac{3.5\sigma}{n^{1/3}} \quad (1)$$

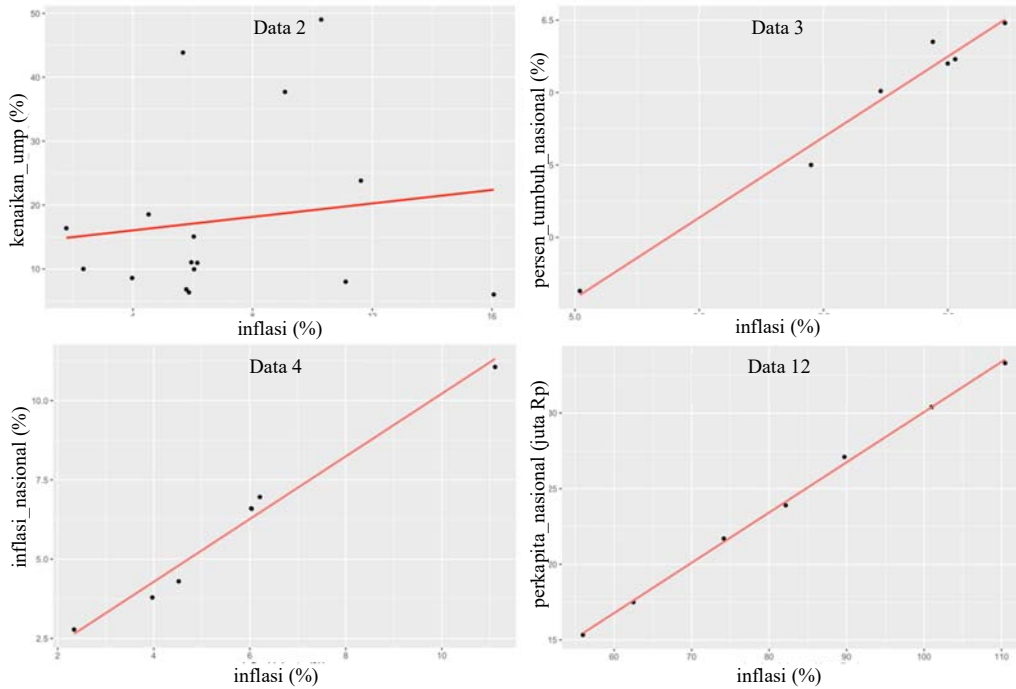
Kelas interval dan frekuensi untuk Data 8, Data 10, dan Data 11 adalah sebagaimana ditunjukkan pada Lampiran Tabel 3, 4, dan 5.

Linear Regression

Linear regression digunakan untuk membuat model linear dua variabel atau lebih, dengan tujuan untuk mencari tahu hubungan antara variabel yang dimodelkan tersebut [9]. Berdasarkan atribut, tipe data, jumlah *field*, dan nilai korelasi *r* [9] yang cukup tinggi yang terdapat pada setiap data, untuk mencari tahu bagaimana *trend* setiap periode maka digunakan metode *linear regression*.

Pada penelitian ini, *linear regression* digunakan untuk Data 2, Data 3, Data 4, dan Data 12. Jumlah variabel yang hendak dimodelkan untuk Data 2, Data 3, Data 4, dan Data 12 berjumlah dua variabel, sehingga model linear yang hendak dibuat yaitu seperti diberikan pada persamaan(2).

$$f(x) = \beta_1x + \beta_0 \quad (2)$$



Gambar 5. Scatter plot dan model linear untuk Data 2, 3, 4, dan 12

β_1 pada persamaan(2) merupakan gradien garis linear, dan β_0 adalah *intercept*. β_1 didapatkan dari persamaan(3).

$$\beta_1 = r \frac{\sigma_y}{\sigma_x} \tag{3}$$

r adalah *correlation coefficient*, yang didapatkan dari persamaan(4) dan σ adalah *sample standard deviation*.

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{(n \sum x^2 - (\sum x)^2)(n \sum y^2 - (\sum y)^2)}} \tag{4}$$

Pemodelan yang dibuat untuk Data 2 adalah memodelkan atribut *kenaikan_ump* dari atribut *inflasi*, untuk Data 3 memodelkan atribut *persen_tumbuh_nasional* dari atribut *persen_tumbuh_jakarta*, untuk Data 4 memodelkan atribut *inflasi_nasional* dari atribut *inflasi_jakarta*, dan untuk Data 12 memodelkan atribut *perkapita_nasional* dari atribut *perkapita_jakarta*. *Scatter plot* pada Data 2, Data 3, Data 4, dan Data 12 adalah sebagaimana ditunjukkan pada Gambar 4.

Untuk plot pada Data 2, terlihat bahwa terdapat *outlier*. *Outlier* tersebut akan dihilangkan untuk membuat model lebih baik.

Dari hasil perhitungan, model linear untuk Data 2 adalah $y = 0.5265x + 13,937$ dengan nilai r sebesar 0.1439, model linear untuk Data 3

adalah $y = 1.263x - 2.062$ dengan nilai r sebesar 0.9078, model linear untuk Data 4 adalah $y = 0.9889x + 0.3296$, dengan nilai r sebesar 0.9892, dan model linear untuk Data 12 adalah $y = 0.3322x - 3.1573$ dengan nilai r sebesar 0.9993. Gambar 5 menunjukkan plot model linear untuk setiap data.

K-Means

K-Means adalah algoritma yang digunakan untuk melakukan pengelompokan data ke dalam bentuk dua atau lebih kelompok. Tujuan pengelompokan adalah meminimalkan variasi di dalam suatu kelompok dan memaksimalkan variasi antarkelompok [10].

Pada *platform* yang memiliki sumber daya terbatas dan untuk meminimalkan waktu latensi, efisiensi sumber daya sangat diperlukan. Untuk mencapai performa terbaik maka diperlukan algoritma *clustering* yang memiliki efisiensi yang cukup baik. Dibandingkan metode *clustering* lain seperti teknik *hierarchical*, *k-Means* memiliki performa $O(n)$ sedangkan teknik *hierarchical* memiliki performa $O(n^2)$ [11].

Data 7 dibagi ke dalam tiga kelompok. *k-Means* digunakan pada Data 7 dengan tujuan untuk meminimalkan variasi setiap kelompok. Data keluaran *k-Means* selanjutnya akan divisualisasikan menggunakan *pie chart* seperti ditunjukkan pada

Gambar 6. Berikut merupakan algoritma *k*-Means [10]:

- 1) Tentukan jumlah kelompok *k*.
- 2) Tentukan titik pusat kluster *k* (*centroid*) secara acak pada ruang *d* dimensi.
- 3) Kelompokkan data sehingga terbentuk *k* buah kluster dengan titik *centroid* dari setiap kluster merupakan titik *centroid* yang telah dipilih sebelumnya.
- 4) Kelompokkan setiap titik ke dalam *k*-set *centroid* paling dekat, yaitu dengan cara menghitung jarak setiap titik dengan pusat kluster. Persamaannya ditunjukkan pada persamaan (5).

$$dist(x_1, x_2) = \sqrt{\sum_{j=1}^p |x_{2j} - x_{1j}|^2} \quad (5)$$

- 5) Perbaharui setiap *centroid* dengan rata-rata (*mean*) titik pada kluster tersebut, yaitu titik-titik yang sudah dikelompokkan sebelumnya.
- 6) Ulang langkah ke-4 dan ke-5 sampai pengelompokan tidak berubah.

Pada Data 7, atribut yang menjadi titik-titik dalam *k*-Means adalah atribut *volume* dan nilai. Inisialisasi nilai *centroid* untuk kluster *k*₁ adalah (301900.47, 3933797.8). Sedangkan untuk *k*₂ dan *k*₃ adalah (42286.14, 102493.1) dan (440518.57, 977354.3) berturut-turut.

Setelah iterasi pertama, didapatkan *k*₁ sebanyak 160 observasi (82.05 %), *k*₂ sebanyak 22 observasi (11.28 %), dan *k*₃ sebanyak 12 observasi (6.15 %). Setelah iterasi ke-2 didapatkan bahwa *k*₁ sebanyak 165, observasi (84.62 %), *k*₂ sebanyak 17 observasi (8.7628 %), dan *k*₃ sebanyak 12 observasi (6.1855 %); Setelah iterasi ke-3 didapatkan bahwa *k*₁ sebanyak 165, observasi (84.62 %), *k*₂

sebanyak 17 observasi (8.7628 %), dan *k*₃ sebanyak 12 observasi (6.15 %). Setelah iterasi ke-2, pengelompokan tidak berubah. Gambar 7 menunjukkan plot *final* hasil pengelompokan *k*-Means. Lampiran Tabel 6 menunjukkan statistik atribut yang dihitung per kluster.

Arsitektur Sistem

Perancangan arsitektur (Gambar 8) menjelaskan bagaimana sistem bekerja secara abstrak. Terdapat dua subsistem yang bekerja, *restfull service* dan subsistem *mobile*. Pada subsistem *mobile*, *request* data akan dikirim ke-*restfull service*, setelah itu *restfull service* akan memeriksa apakah data yang diminta. Jika data tersedia, *restfull service* akan memberikan *response* sesuai data yang diminta dan dikirim dalam bentuk format data json. Selanjutnya subsistem *mobile* akan mengolah data json menjadi visualisasi.

Maintainer

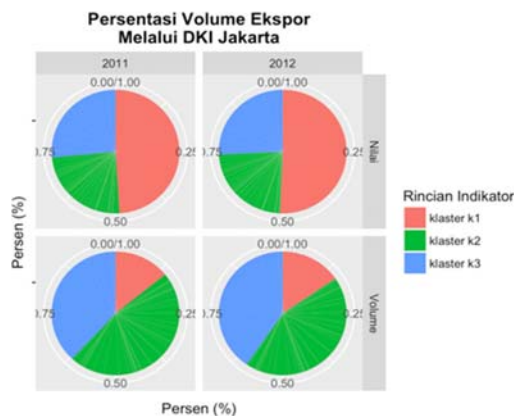
Maintainer adalah pengelola yang mengelola perangkat lunak, menangani *bugs* dan pengembangan perangkat lunak. *Maintainer* dapat melakukan metode POST, GET, PUT, PATCH, dan DELETE melalui protokol HTTP.

Cron Job

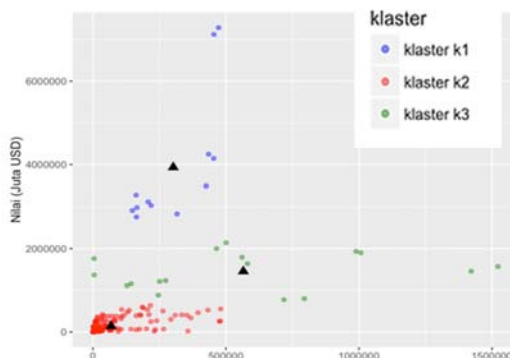
Cron Job adalah sebuah *task scheduler* yang akan memeriksa secara berkala apakah terdapat perubahan pada Jakarta *Open Data* melalui R *Server*. Jika terdapat perubahan, *Cron Job* akan melakukan HTTP *request* yang sesuai ke RESTful *Server*.

R Server

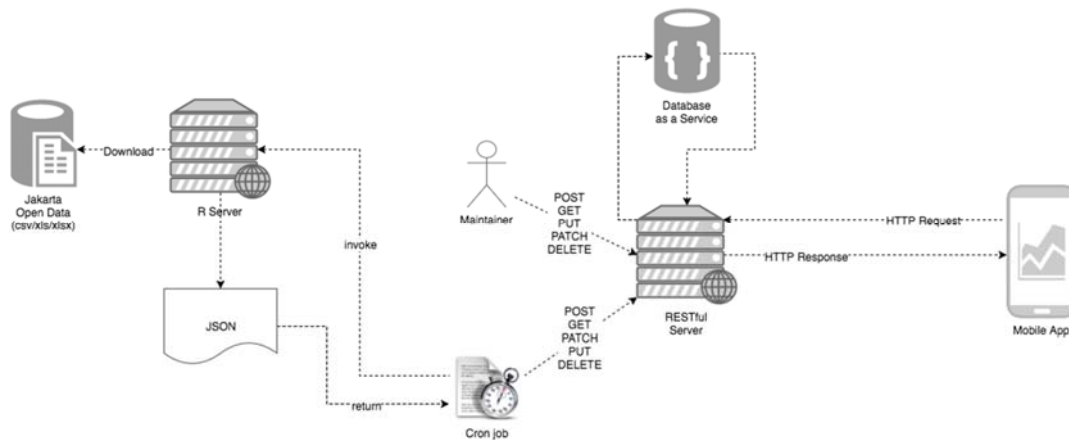
R *Server* adalah *server* yang melakukan analisis terhadap data yang di-*download* dari Jakarta *Open Data*. Pada R *Server* juga algoritma diimplementasikan. Sehingga hasil keluaran dari analisis adalah data dalam format json yang siap di *parse* menjadi visualisasi.



Gambar 6. Pie chart digunakan untuk memudahkan pengguna membandingkan data di setiap kluster



Gambar 7. Plot final keluaran *k*-Means



Gambar 8. Arsitektur sistem

Jakarta Open Data

Jakarta Open Data adalah tempat penyimpanan (*repository*) data terbuka (*open data*) Pemerintah Provinsi DKI Jakarta. Data yang ada pada Jakarta Open Data dalam format *flat file* csv, xls, atau xlsx.

RESTful server

RESTful server adalah server yang meng-host sekaligus menjalankan RESTful API. Data disimpan terpisah pada platform *Database as a Service*.

Database as a Service

Database as a Service adalah server basis data yang digunakan dalam penyimpanan hasil analisis R Server.

Mobile app

Mobile app adalah aplikasi visualisasi yang didesain untuk platform *mobile*, data untuk aplikasi dibaca dari RESTful API melalui protokol HTTP. Gambar Lampiran 1 pada Lampiran adalah contoh beberapa potret visualisasi data pada aplikasi *mobile* yang dibangun.

4. Kesimpulan

Visualisasi interaktif berbasis *mobile* digunakan untuk memudahkan membaca informasi data terbuka Pemprov DKI Jakarta topik ekonomi dan keuangan daerah. Platform *mobile* dipilih karena memiliki keunggulan praktis. Selain itu, platform dibangun di atas *framework* ionic 2, yang mana memiliki keunggulan bisa di-install tidak terbatas pada satu jenis platform *mobile* saja.

Pada *post* kuesioner yang dilakukan pada tanggal 5 Agustus 2016 s/d 10 Agustus 2016 dengan responden yang sama, sebanyak 83.6% responden dapat membaca analisis visual dan mengambil *insight* dari data, sebanyak 14.0% responden dapat membaca analisis visual tetapi tidak dapat

mengambil *insight*, dan sebesar 2.4% responden tidak dapat membaca analisis visual. 2.4% responden yang tidak dapat membaca analisis visual bukan karena visualisasi yang buruk, tetapi lebih karena faktor teknis pada platform yang digunakan.

Kelemahan dari aplikasi visualisasi *mobile* ini terletak pada penanganan pembaharuan data (*error handling*) melalui *script* *corn-job*, di mana *script* *corn-job* hanya bisa menangani data yang tersimpan pada berkas yang sama. Artinya, ketika data yang baru tersedia pada berkas yang berbeda, maka, pembaharuan data harus dilakukan secara manual. Kelemahan lain terletak pada *hybird* platform ionic 2, yang memiliki performa kurang baik untuk beberapa *case* pengguna seperti *crash*.

Visualisasi data bukanlah hal yang baru, berbagai visualisasi data yang memberikan narasi yang bagus terhadap data banyak ditemukan, namun tidak banyak visualisasi data yang baik yang bersumber dari data terbuka (*open data*) Indonesia. Penelitian visualisasi data kedepannya bisa diaplikasikan tidak terbatas menggunakan data topik ekonomi dan keuangan daerah saja.

Referensi

- [1]. C. Snijders, U. Matzat and U.-d. Reips, "Big Data": Big Gaps of Knowledge in the Field of Internet Science," International Journal of Internet Science, vol. 7, no. 1, pp. 1-5, 2013.
- [2]. Anon, "The Open Definition," [Online]. Available: <http://opendefinition.org/>. [Accessed 03 October 2015].
- [3]. I. J. Asmara, E. Achelia, W. Maulana, R. Wijayanti and Y. Rianto, "Teknik Visualisasi Grafik Berbasis Web di Atas Platform Open Source," Seminar Nasional Aplikasi Teknologi Informasi 2009 (SNATI 2009), pp. 44-47, 2009.
- [4]. B. Fry, Visualizing Data, O'Reilly, 2007, pp. 5-14.

- [5]. C. O'Neil and R. Schutt, *Doing Data Science*, O'Reilly, 2013, pp. 34-37.
- [6]. A. Abela, "Chart Suggestions—A thought-starter," in *Extreme Presentation*, 2009.
- [7]. D. C. Montgomery and G. C. Runger, *Applied Statistics and Probability for Engineers Third Edition*, Jhon Wiley & Sons, 2002, pp. 203-206.
- [8]. D. W. Scott, "On Optimal and Data-Based Histograms," *Biometrika*, vol. 66, no. 3, pp. 605-610, 1979.
- [9]. A. M. Glenberg and M. E. Andrzejewski, *Learning from Data, an Introduction to Statistical Reasoning*, 3rd Edition ed., New York: Lawrence Erlbaum Associates, 2008, pp. 477-496.
- [10]. H. Jiawei, M. Kamber and J. Pei, *Data Mining: Concepts and Techniques*, Third Edition, Waltham: Elsevier, 2012, pp. 451-454.
- [11]. M. Steinbach, G. Karypis and V. Kumar, "A Comparison of Document Clustering Techniques," in *KDD workshop on text mining*, 2000.

Lampiran

LAMPIRAN TABEL II
PEMETAAN KEBUTUHAN INFORMASI TERHADAP ALGORITMA/STATISTIK

Kebutuhan Informasi	Data	Pemecahan
Seberapa besar sumber daya telah dimanfaatkan di dalam kegiatan ekonomi.	1, 9	Statistik
Bagaimana <i>trend</i> variabel-variabel ekonomi mengalami pergerakan	2, 3, 4, 12	<i>Linear Regression</i>
Bagaimana distribusi harga suatu komoditas.	8, 10, 11	Histogram
Seberapa besar permintaan atau penawaran terhadap komoditas.	7	<i>k</i> -Means

LAMPIRAN TABEL III
KELAS INTERVAL DAN FREKUENSI DATA 8

Kelas Interval	Frekuensi
0 - 289.93	51
289.94 - 579.87	42
579.88 - 869.81	4
869.82 - 1159.75	6
1159.76 - 1449.69	7
1449.70 - 1739.63	5

LAMPIRAN TABEL V
KELAS INTERVAL DAN FREKUENSI DATA 11

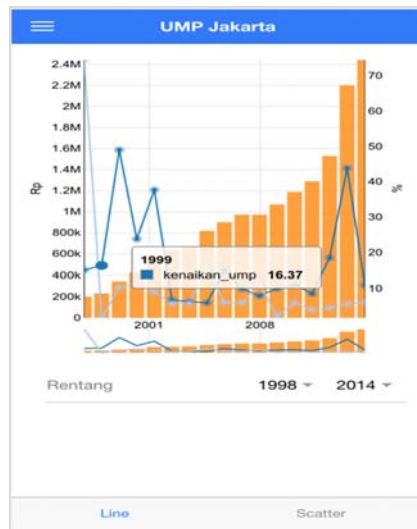
Kelas Interval	Frekuensi
0 - 2825.27	0
2825.28 - 5650.55	0
5650.56 - 8475.83	138
8475.84 - 11301.12	695
11301.13 - 14126.40	297
14126.41 - 16951.68	73
16951.69 - 19776.96	52
19776.97 - 22602.24	40
22602.25 - 25427.52	49
25427.53 - 28252.80	54
28252.81 - 31078.08	62
31078.09 - 33903.36	18
33903.37 - 36728.64	11
36728.65 - 39553.92	8
39553.93 - 42379.20	1
42379.21 - 45204.48	0
45204.49 - 48029.76	4
48029.77 - 50855.04	4
50855.05 - 53680.32	13
53680.33 - 56505.60	16
56505.61 - 59330.88	5

LAMPIRAN TABEL IV
KELAS INTERVAL DAN FREKUENSI DATA 10

Kelas Interval	Frekuensi
0 - 17753.08	105
17753.09 - 35506.17	123
35506.18 - 53259.26	17
53259.27 - 71012.35	0
71012.36 - 88765.44	0
88765.45 - 106518.50	7
106518.51 - 124271.60	21
124271.60 - 142024.70	7

LAMPIRAN TABEL VI
STATISTIK ATRIBUT PER KLASTER

Klaster	Max		Min		Mean	
	volume	nilai	volume	nilai	volume	nilai
Klaster k_1	472300.3	7281767	148100.1	2756790	301900.5	3933798
Klaster k_2	480066.6	634643	20.62	3.85	68058.43	138135.5
Klaster k_3	1522428	2140779	774358.9	4056.31	565182.8	1454812



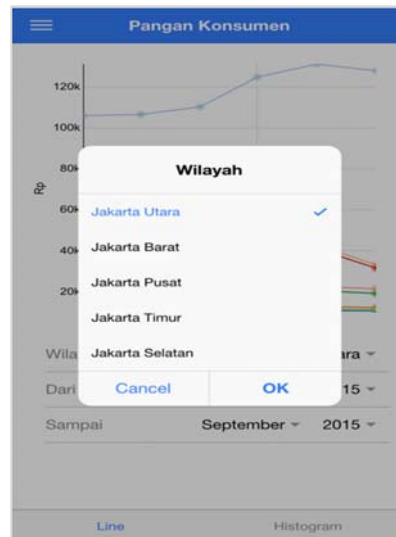
(a)



(b)



(c)



(d)

Lampiran Gambar 1. (a) Visualisasi *line* dan *bar chart* untuk data UMP pada Data 2; (b) Visualisasi *Pie Chart* untuk Data 7; (c) Pengguna dapat melakukan *filtering* dari hasil keluaran *k-Means* yang diaplikasikan pada Data 7; (d) Pilihan yang bisa dilakukan pada *line chart* dari Data 10