

SEGMENTASI TRAFU LISTRIK MENGGUNAKAN ALGORITMA K-MEANS UNTUK MENDUKUNG EVALUASI KAPASITAS GARDU INDUK LISTRIK DI JAWA TIMUR

Rully Agus Hendrawan, Amalia Utamima, Dyanika Ayu Savitri
Jurusan Sistem Informasi, Fakultas Teknologi Informasi, Institut Teknologi Sepuluh Nopember
Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia
e-mail: ruhendrawan@gmail.com

Abstract

The rising in the amount of electricity sold shows that electricity demand in East Java is always increasing. Currently, the increase in peak load is not proportional to the number of installed capacity. This condition will affect in some areas experiencing a shortage of electricity supply at certain times. Enhancement of the effectiveness in distribution of electrical energy to all regions in East Java requires evaluation of the condition of the electrical substation capacity installed. Evaluation was conducted in order to prioritize the allocation of resources and effort to develop the electric substation's capacity. The main electric substations' transformers are grouped based on the conditions and characteristics by using the K-Means algorithm. Group of electric substations' transformer are shown on a map to facilitate the evaluation of the capacity.

Abstrak

Kenaikan jumlah energi listrik yang terjual menunjukkan bahwa kebutuhan listrik di Jawa Timur selalu mengalami peningkatan. Saat ini, peningkatan beban puncak tidak sebanding dengan jumlah kapasitas terpasang. Resikonya beberapa daerah mengalami kekurangan pasokan listrik pada waktu tertentu. Peningkatan efektivitas penyaluran energi listrik ke seluruh daerah di Jawa Timur memerlukan evaluasi terhadap kondisi kapasitas gardu induk listrik yang terpasang. Evaluasi dilakukan supaya dapat memprioritaskan alokasi sumber daya dan tenaga untuk pengembangan kapasitas. Trafo-trafo gardu induk dikelompokkan berdasarkan kondisi dan karakteristiknya dengan menggunakan algoritma K-Means. Kelompok trafo gardu induk listrik ditampilkan pada peta sehingga memudahkan evaluasi kapasitas.

Kata kunci: Clustering, K-Means, Evaluasi Kapasitas

1. PENDAHULUAN

Energi listrik terjual di Jawa Timur sebesar 28,708.11 GWh pada tahun 2013 [1]. Penjualan ini meningkat 6,68% dari tahun sebelumnya. Selain dari peningkatan penjualan listrik, peningkatan kebutuhan akan listrik juga nampak dari Rasio Elektrifikasi di Jawa Timur pada tahun 2013 yaitu sebesar 79,21% [1]. Rasio Elektrifikasi ini menunjukkan bahwa sekitar 20,79% penduduk Jawa Timur masih belum mendapatkan listrik PLN. Hal ini sangat berdampak kepada pasokan listrik nasional mengingat Jawa Timur adalah provinsi dengan jumlah penduduk terbanyak kedua di Indonesia yaitu sebanyak 37.476.757 jiwa [2].

Untuk memenuhi kebutuhan listrik tersebut, dibutuhkan kapasitas terpasang yang lebih dari jumlah penjualan listrik agar kebutuhan listrik dapat selalu terpenuhi. Pada akhir Desember

2013, total kapasitas terpasang dan jumlah unit pembangkit PLN mencapai 34.206 MW dan 4.925 unit, dengan 26.768 MW (78,26%) berada di Jawa [1]. Total kapasitas terpasang meningkat 3,96% dibandingkan dengan akhir Desember 2012. Sedangkan beban puncak pada tahun 2013 mencapai 30.834 MW, meningkat 6,76% dibandingkan tahun sebelumnya [1].

Dapat dilihat bahwa peningkatan jumlah kapasitas terpasang tidak sebanding dengan peningkatan beban puncak yang diketahui dari hasil penjualan listrik. Hal ini dapat menimbulkan beberapa daerah kekurangan pasokan listrik. Jika hal tersebut tidak segera diatasi, dikhawatirkan bahwa daerah yang kekurangan pasokan listrik akan semakin bertambah setiap tahunnya.

Berdasarkan permasalahan diatas, PT. PLN (Persero) Distribusi Jawa Timur harus siap untuk menghadapi permintaan listrik yang

semakin meningkat dan melakukan evaluasi terhadap kapasitas gardu induk listrik yang terpasang dalam menyalurkan energi kebutuhan listrik ke depannya dapat selalu terpenuhi. Jika beban listrik melebihi kapasitas dari trafo dalam gardu induk listrik, maka trafo mengalami panas dan akhirnya meledak yang dapat mengakibatkan kebakaran [3].

Untuk memudahkan proses evaluasi kapasitas listrik, penelitian ini melakukan pengelompokan terhadap kondisi trafo saat ini, analisis kondisi dan karakteristik trafo gardu induk listrik pada setiap kelompok. Hal ini dilakukan agar perusahaan mampu memprioritaskan sumber daya dan tenaganya dalam pengembangan kapasitas.

Batasan dalam penelitian ini adalah pengelompokan yang dilakukan hanya mencangkup pada 201 trafo dalam gardu induk listrik di Jawa Timur, data yang digunakan adalah data beban trafo pada bulan Januari 2014 hingga Desember 2014, besar beban pada tiap trafo ditunjukkan oleh besar kapasitas terpakai yang disebut dengan utilitas dan dituliskan dalam satuan persen, dan analisis kondisi trafo tidak mencangkup material dan sumber daya trafo.

2. METODOLOGI

Penelitian ini dilakukan dengan mengadopsi metodologi data mining Teradata [4]. Terdapat 5 tahap yaitu:

1. Definisi masalah bisnis
2. Persiapan arsitektur dan teknologi
3. Persiapan data
4. Pengembangan, pengujian dan validasi model
5. Pengembangan & penyampaian pengetahuan

Tahapan pertama yaitu definisi masalah bisnis dilakukan menggunakan teknik wawancara dengan beberapa staff pada bagian Perencanaan Sistem Kelistrikan dan pada bagian APD di PT. PLN (Persero) Distribusi Jawa Timur. Pada tahapan ini dirumuskan permasalahan bisnis yang utama adalah bagaimana karakteristik trafo gardu induk yang ada saat ini dan persebarannya. Informasi ini dibutuhkan sebagai pendukung keputusan pada saat memprioritaskan alokasi pengembangan kapasitas pasokan listrik.

Berikutnya pada tahapan persiapan arsitektur dan teknologi. Pada penelitian ini analisis dilakukan terpisah dari sistem perusahaan, sehingga tidak banyak yang perlu dipertimbangkan. Tools yang akan digunakan

untuk menerapkan teknik data mining adalah SPSS dan R [5].

Pada tahapan ketiga yaitu persiapan data. Data utama adalah data gardu induk listrik beserta kapasitas dan beban listrik harian pada Januari 2014 hingga Desember 2014 dalam bentuk arsip dalam format Microsoft Excel. Uji korelasi dilakukan untuk menentukan atribut utama yang menjadi karakteristik trafo.

Tahapan keempat adalah pengembangan dan pengujian model analisis yang dilakukan menggunakan analisis kluster. Terdapat dua metode pada analisis kluster yaitu hirarkikal dan non-hirarkikal. Analisis hirarkikal adalah metode analisis dimana kluster-kluster kecil yang sejenis bergabung secara bertingkat/berjenjang dalam kluster yang lebih besar. Kelemahannya dibandingkan analisis non-hirarkikal adalah komputasi yang lebih lambat. Pada penelitian ini analisis hirarkikal belum memiliki peran yang signifikan karena hanya digunakan sebagai pembandingan jumlah kluster yang ideal.

Analisis kluster hirarkikal dilakukan dengan metode *agglomerative* dilakukan dengan algoritma Ward [6] dan menggunakan perhitungan jarak Euclidean yang cocok untuk atribut yang bertipe kontinyu. Pendekatan *agglomerative* dilakukan secara bottom-up, yaitu: 1) mencari dua anggota kelompok yang paling berdekatan, 2) kelompokkan keduanya menjadi satu, 3) cari anggota paling berdekatan berikutnya. Jarak antar anggota dihitung dengan jarak Euclidean (rumus 1).

$$distance(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \dots \dots \dots (1)$$

Pada algoritma Ward, penggabungan kluster dihitung dengan fungsi objektif error sum of square (ESS, rumus 2).

$$ESS = \sum_{i=1}^n x_i^2 - \frac{1}{n} (\sum_{i=1}^n x_i)^2 \dots \dots \dots (2)$$

Pendekatan *agglomerative* menghasilkan tabel *Agglomeration Schedule* dimana nilai dari jarak yang digunakan untuk membentuk kluster (euclidean) ditunjukkan pada kolom koefisien. Nilai koefisien menunjukkan kualitas dari kluster yang dibentuk, semakin kecil berarti anggota kluster semakin homogen / mirip. Pada penelitian ini kluster dibentuk berdasarkan kemiripan, sehingga semakin kecil nilai koefisien maka semakin baik. Jumlah kluster yang ideal didapatkan dengan mencari loncatan terbesar yang pertama dari nilai koefisien (*elbow rule*) pada *Agglomeration Schedule*. [7]

Teknik analisis utama dilakukan menggunakan metode non-hirarkikal dengan algoritma K-Means [8]. Penerapan algoritma K-Means terdiri atas empat tahap [9], yaitu: 1) penentuan lokasi centroid awal sejumlah k secara acak; 2) menghitung jarak antara centroid dengan setiap data; 3) mengelompokkan data berdasarkan jarak terdekat; 4) jika semua anggota kelompok sama seperti perulangan sebelumnya maka berhenti. Jika tidak, maka nilai centroid baru dihitung dari rata2 posisi anggota kelompok dan kembali ke langkah kedua.

Validasi terhadap hasil pengelompokan dilakukan dengan menggunakan Dunn Index [10]. Dari proses validasi ini dapat diketahui kualitas dari klaster. Klaster C yang berjumlah m dinyatakan baik apabila terpisah dengan baik dan padat (rumus 3). Klaster C_i dan C_j terpisah dengan baik apabila jarak minimum antar kedua klaster $d(x,y)$ bernilai besar (rumus 4). Klaster C_i dinyatakan padat apabila varian antar anggota dalam klaster (Δ_k) bernilai kecil. Rumus 5 digunakan untuk menghitung diameter klaster, yaitu nilai terjauh (max) dari seluruh $d(x, y)$ yang merupakan jarak antar anggota dalam klaster.

$$DI_m = \frac{\min_{1 \leq i \leq j \leq m} dist(C_i, C_j)}{\max_{1 \leq k \leq m} diam(C_k)} \dots\dots\dots(3)$$

$$dist(C_i, C_j) = \min_{x \in C_i, y \in C_j} d(x, y) \dots\dots\dots(4)$$

$$diam(C_k) = \max_{x, y \in C_k} d(x, y) \dots\dots\dots(5)$$

Tahap kelima adalah tahapan analisis hasil pengelompokan. Pada tahap ini dilakukan analisis karakteristik beban listrik pada masing-masing kelompok yaitu kondisi dan karakteristik masing-masing trafo gardu induk listrik berdasarkan zona waktu, kapasitas, utilitas, bulan, dan Area Pelayanan dan Jaringan (APJ) trafo. Dengan karakteristik ini dapat diketahui trafo gardu induk listrik mana yang dapat diprioritaskan oleh perusahaan dalam pengembangannya. Kelompok trafo gardu induk divisualisasikan pada peta Jawa Timur untuk memudahkan evaluasi kapasitas.

3. HASIL DAN DISKUSI

Atribut Penting Karakteristik Trafo

Data utama berisikan atribut antara lain nama gardu induk, nomor trafo, Area Pengatur Distribusi (APD), Area Pelayanan dan Jaringan (APJ), kapasitas, beban puncak harian siang, dan beban puncak harian malam. Dari beberapa atribut tersebut dilakukan uji korelasi untuk

menentukan atribut yang penting bagi perusahaan sebagai pembentuk karakteristik trafo. Hasil uji korelasi menunjukkan bahwa atribut atribut zona waktu, kapasitas, dan utilitas trafo memiliki tingkat relevansi yang tinggi sehingga nilainya saling mempengaruhi antar atribut satu dengan yang lain. Data yang digunakan untuk tahapan selanjutnya terdiri atas atribut nama_trafo, waktu, kapasitas, dan utilitas.

Penentuan Jumlah Cluster

Menggunakan Metode Ward

Metode Ward merupakan metode yang digunakan dalam penelitian ini untuk membantu menentukan jumlah *cluster* optimal. Hasil yang didapatkan adalah tabel *Agglomeration Schedule* yang dapat dilihat pada tabel 1.

Tabel 9. Cuplikan Hasil *Agglomeration Schedule*

Stage	Clusters	Coefficients	Coefficients Difference
190	10	48.500	5.173
191	9	53.738	5.238
192	8	60.157	6.418
193	7	73.332	13.175
194	6	89.005	15.672
195	5	112.277	23.273
196	4	139.387	27.110
197	3	203.471	64.083
198	2	272.065	68.594
199	1	378.719	106.655
200	0	599.999	221.280

Jumlah *cluster* yang optimal didapatkan dengan melihat loncatan nilai koefisien terbesar yang pertama. Pada tabel 5.1 terlihat bahwa loncatan koefisien antara *stage* 196 dengan 197 adalah sebesar 64.083 dan merupakan loncatan terbesar pertama diantara *stage* lainnya. Maka dapat ditarik kesimpulan bahwa jumlah *cluster* yang optimal untuk digunakan pada *clustering* dengan algoritma *k-means* adalah sejumlah 4 *cluster*.

Proses Clustering Menggunakan

Algoritma K-Means

Jumlah *cluster* yang optimal menurut metode Ward adalah 4 yang akan digunakan sebagai nilai parameter k pada K-means. Hasil yang didapatkan adalah nomor klaster untuk setiap trafo dan jaraknya dari centroid. Sebagian hasil dari proses *clustering* ini dapat dilihat pada tabel 2.

Tabel 10. Cuplikan Hasil Clustering Trafo

Case Number	Trafo	Cluster	Distance
1	ALTA PRIMA1	3	1.053
2	ALTA PRIMA2	2	.266
3	BABADAN1	3	.668
4	BABADAN2	3	.649
5	BABAT1	1	.330
6	BABAT2	3	.968
7	BALONG BENDO1	3	1.239
8	BALONG BENDO3	3	1.059
9	BANARAN4	1	.269
10	BANARAN5	4	1.372

Selanjutnya dihitung jarak antara titik pusat satu *cluster* dengan titik pusat *cluster* lainnya dari hasil *final cluster center*. Dengan mengevaluasi jarak dalam tabel 3, dapat disimpulkan bahwa kedekatan antar *cluster* yang terbentuk adalah *cluster* 1-4-3-2. Informasi jarak ini dapat digunakan untuk membantu dalam menentukan keputusan tingkat kepentingan trafo yang akan dilakukan bersama pihak perusahaan.

Tabel 11. Distance Between Final Cluster Center

Cluster	1	2	3	4
1		3.404	2.267	1.915
2	3.404		2.156	2.184
3	2.267	2.156		2.210
4	1.915	2.184	2.210	

Jumlah kasus atau anggota yang terdapat pada masing-masing *cluster* yang terbentuk dapat dilihat pada tabel 4. *Cluster* 1 terdiri dari 60 trafo, *cluster* 2 terdiri dari 33 trafo, *cluster* 3 terdiri dari 63 trafo, dan pada *cluster* 4 terdiri dari 45 trafo. *Cluster* 3 memiliki jumlah trafo paling banyak dari total trafo sebanyak 201 trafo. Kesimpulan dari hasil pada tabel ini adalah urutan *cluster* dari yang memiliki anggota paling banyak hingga paling sedikit adalah *cluster* 3-1-4-2.

Tabel 12. Number of Cases in Each Cluster

Cluster	Number of Cases
1	60
2	33
3	63
4	45

Valid	201
Missing	

Validasi Hasil Clustering dengan Dunn Index

Dunn Index digunakan sebagai metode validasi untuk melihat apakah *cluster* yang didapatkan telah optimal. Pada penelitian ini jumlah *cluster* yang dimaksud adalah nilai k yang digunakan dalam proses *clustering* menggunakan algoritma K-means.

Tabel 13. Nilai Validasi Dunn Index

Jumlah Cluster	Dunn Index (DI)
3	0.04776848
4	0.05721716
5	0.03186556
6	0.03869601
7	0.03869601
8	0.03996762

Uji coba dilakukan dengan nilai k mulai dari 3 hingga 8. Hasil perhitungan Dunn Index dapat dilihat pada tabel 5. Nilai DI untuk jumlah *cluster* 4 merupakan nilai yang paling tinggi dibandingkan nilai yang lain. Berdasarkan pada teori yang dijelaskan oleh Halkidi, dkk [11], nilai DI yang tinggi menunjukkan bahwa *cluster* telah terpisah dengan baik dan padat. Meskipun nilai DI *cluster* 4 ini di bawah 1, namun nilai ini lebih tinggi dibandingkan nilai *cluster* lainnya. Sehingga dapat disimpulkan bahwa jumlah *cluster* 4 adalah jumlah yang optimal.

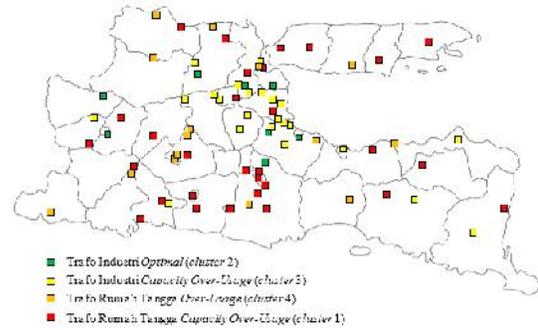
Analisis Karakteristik Trafo

Setelah melakukan analisis hasil *clustering* berdasarkan atribut-atribut yang digunakan maka dapat diketahui karakteristik dari trafo pada masing-masing *cluster*. Hasil analisis karakteristik trafo dapat dilihat pada tabel 6 berikut.

Tabel 14. Analisis Karakteristik Trafo

<u>Cluster 1</u> Trafo Rumah Tangga Capacity Over-Usage
- Utilitas tertinggi trafo berada pada zona waktu malam hari
- Trafo berkapasitas 6-30 MVA
- Utilitas trafo mayoritas telah mencapai lebih dari 80% dengan utilitas tertinggi antara lain:
1. 107.27% pada trafo Tuban1
2. 102.01% pada trafo Porong1
3. 99.59% pada trafo Probolinggo1
- Trafo pada <i>cluster</i> 1 cenderung mengalami peningkatan kebutuhan listrik pada akhir tahun (November-Desember)
- Mayoritas trafo terletak di Area Pelayanan

dan Jaringan (APJ) Malang dan Kediri
Cluster 4 Trafo Rumah Tangga Over-Usage
<ul style="list-style-type: none"> - Utilitas tertinggi trafo berada pada zona waktu malam hari - Sebagian besar trafo berkapasitas 60 MVA - Utilitas beberapa trafo telah mencapai lebih dari 80% dengan utilitas tertinggi antara lain: <ol style="list-style-type: none"> 1. 96.16% pada trafo Kenjeran1 2. 96.13% pada trafo Kertosono2 3. 95.84% pada trafo Banaran5 - Trafo pada <i>cluster</i> 4 cenderung mengalami peningkatan kebutuhan listrik pada akhir tahun (Oktober-Desember) - Mayoritas trafo terletak di Area Pelayanan dan Jaringan (APJ) Surabaya Selatan dan Surabaya Utara
Cluster 3 Trafo Industri Capacity Over-Usage
<ul style="list-style-type: none"> - Utilitas tertinggi trafo berada pada zona waktu siang hari - Trafo berkapasitas 10-60 MVA - Utilitas trafo mayoritas telah mencapai 80% dengan utilitas tertinggi antara lain: <ol style="list-style-type: none"> 1. 99.88% pada trafo Mojokerto6 2. 99.42% pada trafo Mranggen2 3. 99.30% pada trafo Ploso2 - Trafo pada <i>cluster</i> 3 cenderung mengalami peningkatan kebutuhan listrik pada pertengahan awal tahun (Maret-Juni) - Mayoritas trafo terletak di Area Pelayanan dan Jaringan (APJ) Pasuruan dan Sidoarjo
Cluster 2 Trafo Industri Optimal
<ul style="list-style-type: none"> - Utilitas tertinggi trafo paling banyak berada pada zona waktu siang hari - Trafo berkapasitas 60 MVA - Utilitas trafo mayoritas mencapai 40%-80% dengan utilitas tertinggi antara lain: <ol style="list-style-type: none"> 1. 66.51% pada trafo Kupang1 2. 64.66% pada trafo Tuban2 3. 62.70% pada trafo Waru3 - Trafo pada <i>cluster</i> 2 mengalami peningkatan kebutuhan listrik pada akhir tahun (Desember) - Mayoritas trafo terletak di Area Pelayanan dan Jaringan (APJ) Surabaya Utara dan Pasuruan



Gambar 1. Peta Persebaran Gardu Induk per Cluster di Jawa Timur

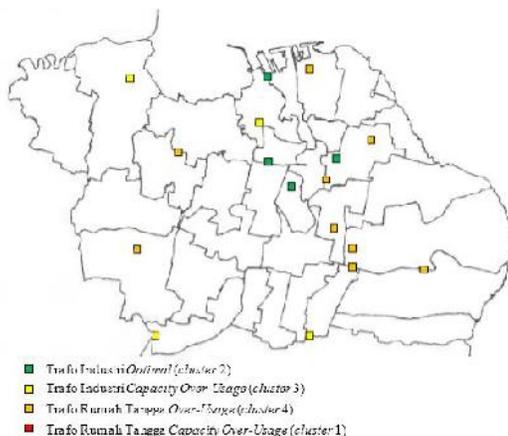
Hasil keempat *cluster* tersebut dibuat dalam bentuk peta dengan titik persebaran gardu induk listrik di Jawa Timur. Persebaran trafo hasil *clustering* tersebut dapat dilihat pada gambar 2 berikut.

Tiap titik gardu induk listrik pada gambar 2 tersebut mewakili trafo-trafo yang berada dalam satu gardu induk listrik. Kemudian setiap titik diberi warna sesuai dengan *cluster* dari trafo pada gardu induk listrik tersebut. Jika trafo termasuk dalam *cluster* 1 maka diberi tanda titik dengan warna merah, jika trafo termasuk dalam *cluster* 2 maka titik tersebut diberi warna hijau, jika trafo termasuk dalam *cluster* 3 maka diberi warna dengan kuning, dan yang terakhir jika trafo termasuk ke dalam *cluster* 4 maka titik diberi dengan warna jingga.

Pada gambar 2 menunjukkan titik-titik lokasi seluruh gardu induk listrik di Jawa Timur kecuali gardu induk listrik yang berada di dalam wilayah Surabaya. Persebaran gardu induk listrik di Surabaya ditunjukkan dengan peta wilayah Surabaya pada gambar 3. Persebaran gardu induk listrik di kota Surabaya dibuat gambar peta khusus kota Surabaya karena pada kota Surabaya sendiri terdapat 17 gardu induk listrik. Jika gambar lokasi gardu induk listrik di Surabaya menjadi satu dengan gambar lokasi gardu induk listrik di Jawa Timur maka persebarannya tidak dapat terlihat dengan jelas karena ke-17 titik gardu induk listrik akan berkumpul dalam satu titik saja.

Persebaran Trafo Hasil Clustering

Pembuatan peta pada penelitian ini bertujuan agar memudahkan pengamatan dan pemahaman terhadap lokasi hasil *clustering*. Dari 201 trafo gardu induk listrik didapatkan 4 *cluster* sesuai dengan atribut zona waktu, kapasitas, dan utilitas yang digunakan dalam proses *clustering*.



Gambar 2. Peta Persebaran Gardu Induk per Cluster di Kota Surabaya

4. KESIMPULAN

Beberapa kesimpulan yang dapat diambil dari penelitian ini diantaranya adalah sebagai berikut:

1. Dengan uji korelasi, didapatkan bahwa atribut yang digunakan sebagai kriteria penentuan trafo paling penting bagi perusahaan adalah zona waktu, kapasitas, dan utilitas trafo.
2. Berdasarkan hasil *clustering* dengan metode Ward dan K-Means jumlah pengelompokan trafo gardu induk listrik di Jawa Timur adalah 4 *cluster* dari keseluruhan jumlah trafo sebanyak 201. *Cluster* 1 terdiri atas 60 trafo, *cluster* 2 terdiri atas 33 trafo, *cluster* 3 terdiri atas 63 trafo dan *cluster* 4 terdiri atas 45 trafo.
3. Hasil *clustering* divalidasi dengan nilai Dunn Index sebesar 0.05721716 yang menunjukkan bahwa klaster dengan jumlah 4 tersebut terpisah paling baik dan paling padat.
4. Dari hasil analisis karakteristik trafo berdasarkan acuan perusahaan didapatkan bahwa trafo pada *cluster* 2 yang diberi label *Trafo industri optimal* merupakan trafo yang paling penting bagi perusahaan karena kebutuhan listrik yang tinggi dan konstan dilihat dari kondisi kapasitas yang mencapai maksimal yaitu 60 MVA dan utilitas yang tidak melebihi 80% sehingga kelebihan pemakaian trafo pada *cluster* ini harus sangat dihindari.

DAFTAR PUSTAKA

[1] P. P. (Persero), Statistik PLN 2013, Jakarta: Sekretariat Perusahaan PT. PLN (Persero), 2013.
 [2] B. P. S. J. Timur, "Statistik Daerah Provinsi Jawa

Timur," 2013. [Online]. Available: "http://jatim.bps.go.id/index.php?hal=publikasi_detil&id=2".

[3] E. F. Kusuma, "Kebakaran Gardu Induk PLN Cawang Diduga Karena Kelebihan Beban Listrik," detikNews, 3 Oktober 2013. [Online]. Available: "http://news.detik.com/read/2013/10/03/042844/2376182/10/kebakaran-gardu-induk-pln-cawang-diduga-karena-kelebihan-beban-listrik?9911012". [Accessed 10 Februari 2015].

[4] A. Zaima, J. Kashner. (2003), "Data Mining Primer for the Data Warehouse Professional", Business Intelligence Journal, 44–54.

[5] R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

[6] Ward, J. H., Jr. (1963), "Hierarchical Grouping to Optimize an Objective Function", Journal of the American Statistical Association, 58, 236–244.

[7] Marija J. Norušis. (2011), "IBM SPSS Statistics 19 Statistical Procedures Companion", Prentice Hall

[8] Kanungo, Tapas, et al. (2002), "An efficient k-means clustering algorithm: Analysis and implementation." Pattern Analysis and Machine Intelligence, IEEE Transactions on 24.7: 881–892.

[9] Santosa, B. (2007), "Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis". Yogyakarta: Graha Ilmu.

[10] Dunn, J. C. (1973), "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters". Journal of Cybernetics 3 (3): 32–57.

[11] Halkidi M., Batistakis Y., Vazirgiannis M. (2002), "Clustering Validity Checking Methods: Part II," *Clustering algorithms and validity measures*, p. 2.