ANALISIS CLUSTERING PADA DATA PELANGGAN PLN APJ BANDUNG

Oleh

BERRY ZULKARNAEN

10108123



Penguji 1 : Adam Mukharil Bachtiar, S.Kom

Penguji 2 : Dr.Janivita Joto Sudirham, S.T.,MSc

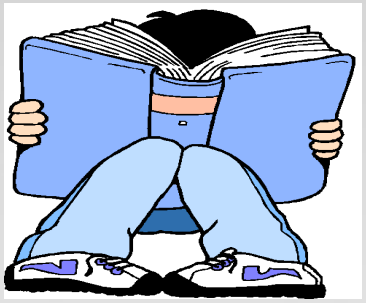
Penguji 3 : Mira Kania Sabariah, S.T.,M.T.

fppt.com

Konten



Pendahuluan



Studi Kasus



Analisis Data



Pemodelan Data





Demo Program

fppt.com



Pendahuluan



fppt.com

Data Mining ?



Prof. Jiawei Han: "Simply stated, data mining refers to *extracting or "mining" knowledge from large amounts of data.* " (Han & Kember,**Data Mining: Concepts and Techniques**)

fppt.com

Pengertian Data Mining

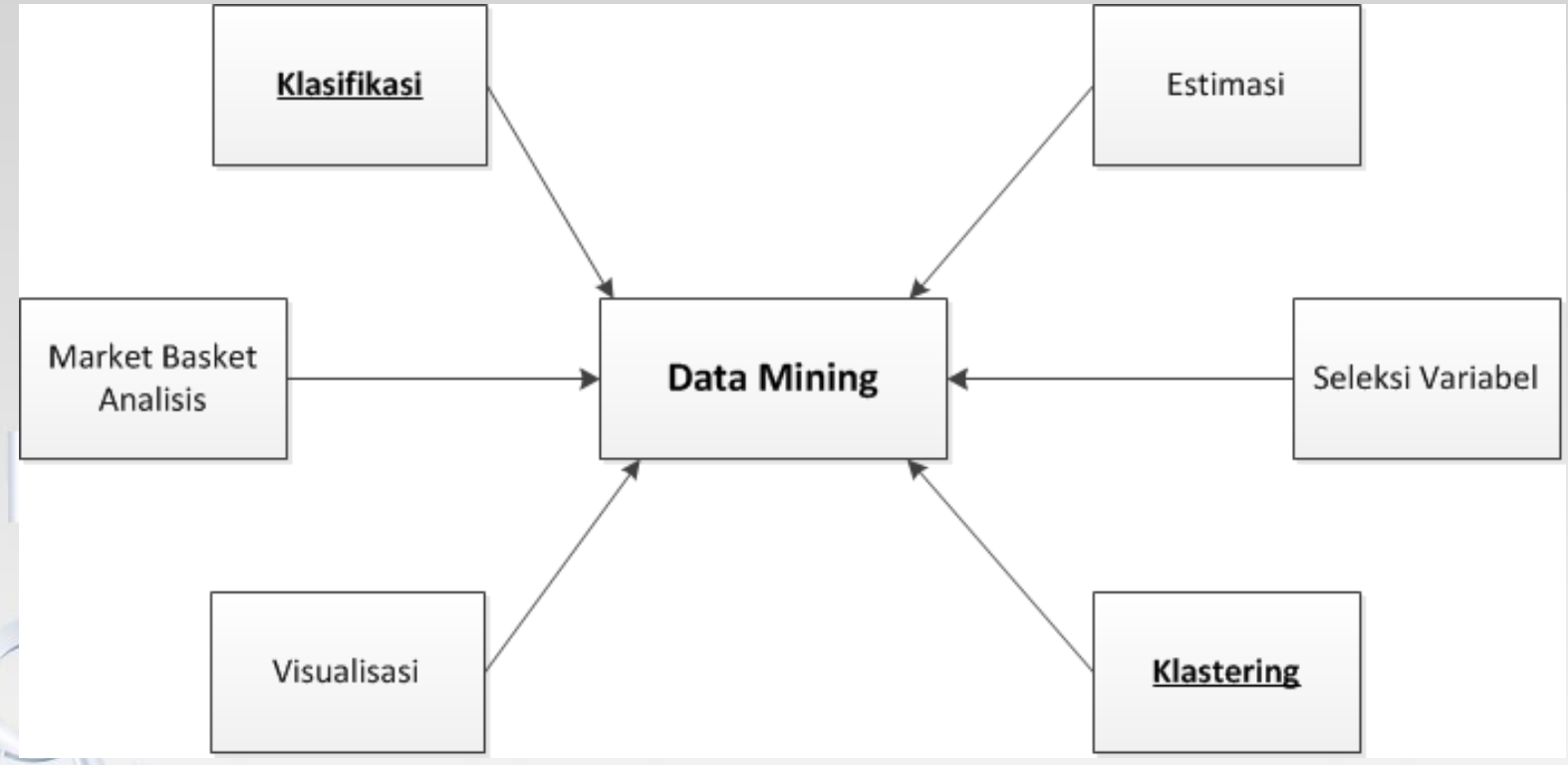
Data Mining, disebut juga KDD (*Knowledge Discovery in Database*) adalah kegiatan yang meliputi pengumpulan, pemakaian **data historis** untuk menemukan **keteraturan**, **pola** atau hubungan dalam ***set data berukuran besar***.



(**Budi Santosa**, Teknik Pemanfaatan Data Untuk Analisis)

fppt.com

Teknik Data Mining



fppt.com

Clustering ?

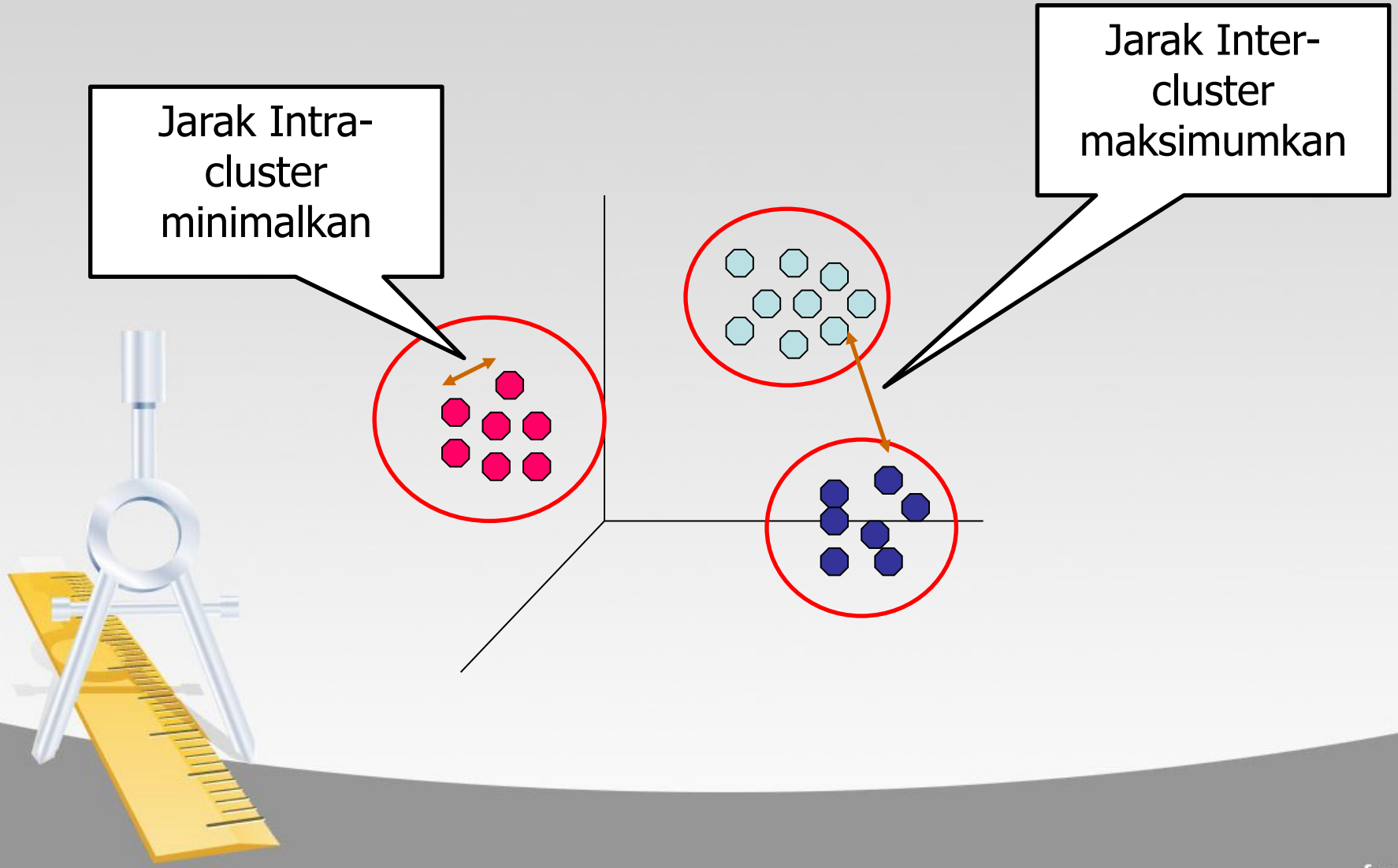
" The process of grouping a set of physical or abstract objects into classes of ***similar*** objects "

(**Han & Kember**, Data Mining: Concepts and Techniques,Ch.383)

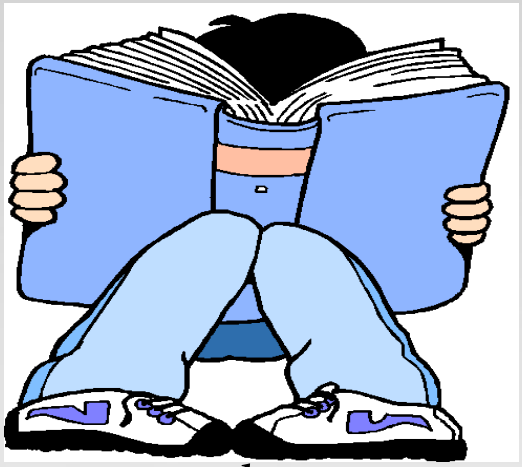


fppt.com

Contoh Clustering



fppt.com



Studi Kasus



fppt.com

PT PLN (Persero) Distribusi Jawa Barat dan Banten

APJ BANDUNG

"Electricity for a Better Life"

www.pln.co.id

fppt.com

Latar Belakang

1. Berangkat dari **perbedaan** akan kebutuhan energi listrik setiap pelanggan.
2. *Penggolongan tarif* pemakaian listrik saat ini apakah masih efektif.
3. Pola persebaran energi listrik berbeda di setiap **wilayah**.



fppt.com

Perumusan Masalah

Bagaimana menganalisis **pengelompokan** pelanggan dalam pemakaian energi listrik di wilayah kota bandung



fppt.com

Maksud

Menerapkan teknik klastering untuk **membentuk kelompok** berdasarkan jumlah pemakaian listrik



Tujuan

1. Mampu memperbaiki penggolongan pelanggan berdasarkan analisis yang telah dilakukan
2. Mampu memberikan gambaran tentang jumlah pemakaian listrik di setiap wilayah kota Bandung.
3. Membantu pihak perusahaan dalam pengambilan keputusan.

fppt.com

Batasan Masalah(1)

1. Pengelompokan pelanggan untuk menemukan pola perilaku pemakaian listrik dari bulan ke bulan.
2. Menganalisis data pelanggan berdasarkan pemakaian listrik (Kwh).
3. Metode yang digunakan adalah Teknik Klastering



fppt.com

Batasan Masalah(2)

1. Metode clustering yang digunakan adalah ***K-mean***
2. Aplikasi yang dibangun merupakan implementasi dari algoritma *K-mean.*
3. Menggunakan intepreter *PHP*, *web server WAMP*, dan *DBMS MySQL.*



fppt.com

Metode K-mean ?

K-Means merupakan salah satu metode data clustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster/kelompok.

Sumber : **Yudi Agusta, PhD** :*Jurnal Sistem dan Informatika Vol. 3 (Pebruari 2007), 47-60,*



fppt.com

K-means Untuk Pengelompokan

Metode ini **mempartisi** data ke dalam cluster/kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain.



Sumber : **Yudi Agusta, PhD** :*Jurnal Sistem dan Informatika Vol. 3 (Pebruari 2007), 47-60,*

fppt.com

Algoritma K-mean

1. Tentukan jumlah *cluster*
2. Alokasikan data ke dalam *cluster* secara random
3. Hitung *centroid*/rata-rata dari data yang ada di masing-masing *cluster*.
4. Alokasikan masing-masing data ke centroid/rata-rata terdekat
5. Kembali ke Step 3, apabila masih ada data yang berpindah *cluster* atau apabila perubahan nilai centroid.

Gambar Flowchart Algoritma K-mean

fppt.com



Analisis Data

fppt.com

Proses ETL

Ekstrak Data

Transform (***Cleaning***)



Loading...

fppt.com

Ekstrak

Mengekstrak data dari sumber data

Data Rekening Pelanggan



fppt.com

Cleaning

Membersihkan Atribut yang tidak digunakan

Data Setelah Proses Cleaning



fppt.com

Loading...

Memuat data untuk selanjutnya dilakukan proses klastering



fppt.com

Contoh Klastering Data Pemakaian Listrik(1)

Dari UPJ Kopo sebanyak 10 *record* dari bulan Desember 2011 s.d April 2012, lalu dilakukan perhitungan rata-rata Kwh pemakaian selama 5 bulan :

Jumlah klaster : 3

Jumlah data :10

Jumlah atribut : 2



fppt.com

Contoh Klastering Data Pemakaian Listrik(2)

No	Alamat	Rata-rata pemakaian Kwh selama 5 bulan
1	DINAS KESEHATAN	452
2	PL KES BOJONGLOA	375.6
3	PL KES BOJONGLOA KM 6	427.2
4	RUMAH POTONG HEWAN	1201.2
5	KANT LINGK CIPARAY	1970.8
6	KANTOR PUSKESMAS	166.8
7	POMPA AIR PDAM	4503.2
8	PES LALULINTAS BDG	2893.4
9	KANTOR DESA	244.6
10	SEKTA BBK CIPARAY	571.8

fppt.com

Contoh Klastering Data Pemakaian Listrik(3)

Jumlah klaster = 3, alokasikan data kedalam klaster secara *random* lalu tentukan nilai *centroid/mean* pada masing-masing klaster seperti pada tabel berikut :

$$Mean = \frac{data_1 + data_2 + data_3 + data_4}{\text{Jml. Data}}$$



fppt.com

Contoh Klastering Data Pemakaian Listrik(3)

Plotting Data Secara Random

C0	C1	C2
Rumah Potong Hewan (1,201)	Kantor Puskesmas (167)	Kant. Lingkungan Ciparay (1,970)
Kantor Desa (245)	PES Lalu Lintas BDG (2,893)	PL KES Bojongloa (376)
PL KES Bojongloa KM 6 (427)	Pompa Air PDAM (4,503)	Dinas Kesehatan (452)
Sekta Ciparay (572)		
Centroid		
611.15	2,520.93	932.53

fppt.com

Contoh Klastering Data Pemakaian Listrik(4)

Alokasikan masing-masing data ke *centroid*/rata-rata terdekat dengan menggunakan rumus *Euclidian distance* d = $\sqrt{(x-y)^2}$, dari data ke-1 sampai data ke-10.

Jarak antara data pertama dengan centroid pertama (C_0)

$$d_{1,0} = \sqrt{(611{,}15 - 1201)^2} = 589{,}85$$

Jarak antara data pertama dengan centroid kedua (C_1)

$$d_{1,1} = \sqrt{(2520{,}93 - 1201)^2} = 1319{,}93$$

Jarak antara data pertama dengan centroid ketiga (C_2)

$$d_{1,2} = \sqrt{(932{,}53 - 1201)^2} = 286{,}47\ldots\ldots\ldots$$

hitung sampai $d_{10,2}$

fppt.com

Contoh Klastering Data Pemakaian Listrik(5)

Setelah dilakukan Iterasi sampai tidak ada data yang berpindah/nilai centroid yang sama dengan Iterasi sebelumnya, maka klastering dianggap sudah optimal



fppt.com

C0	C1	C2
		RUMAH POTONG HEWAN (1,201)
KANTOR PUSKESMAS (167)		
		KANT LINGK CIPARAY (1,970)
Kantor Desa (245)		
	PES LALULINTAS BDG (2,893)	
PL KES BOJONGLOA (376)		
PL KES BOJONGLOA KM 6 (427)		
	POMPA AIR PDAM (4,503)	
DINAS KESEHATAN (452)		
SEKTA BBK CIPARAY (572)		
Centroid		
373	3,698	1,585.5

Klaster Optimal

fppt.com



Pemodelan Data



fppt.com

ERD

Gambar ERD



fppt.com

Diagram Konteks

Gambar Diagram Konteks



fppt.com

Data Flow Diagram Level 1

Gambar DFD Level 1



fppt.com

Data Flow Diagram Level 2 (*Breakdown*)

Gambar DFD Level 2



fppt.com



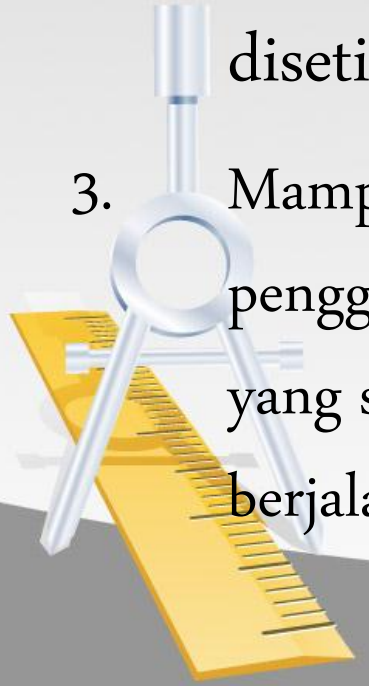
Demo Program



fppt.com

Kesimpulan

1. Dengan dilakukannya analisis data pelanggan menggunakan teknik klastering *data mining*, pihak PLN dapat mengetahui pola perilaku pemakaian listrik.
2. Pihak PLN bisa mengetahui pola pemakaian listrik disetiap wilayah kota Bandung.
3. Mampu memberikan saran serta masukan tentang penggolongan pelanggan, apakah penggolongan pelanggan yang sudah ada sebelumnya masih tepat sasaran sering dengan berjalannya waktu serta kamajuan teknologi.



fppt.com

Saran

1. Diharapkan analisis ini disertai dengan adanya sistem pendukung keputusan.
2. Bisa menghasilkan perangkat lunak yang lebih kompleks tentang sistem pengelompokan.
3. Diharapkan ada metode atau algoritma lain untuk menganalisis sistem pengelompokkan



fppt.com

Terima Kasih



fppt.com