

**PENERAPAN ALGORITMA *K-MEANS* UNTUK
CLUSTERING DATA ANGGARAN PENDAPATAN
BELANJA DAERAH DI KABUPATEN XYZ**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat memperoleh Gelar Sarjana pada Program Studi
Sistem Informasi



Dwi Noviati Nango

531408022

**JURUSAN TEKNIK INFORMATIKA
PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK
UNIVERSITAS NEGERI GORONTALO
JULI 2012**

Intisari

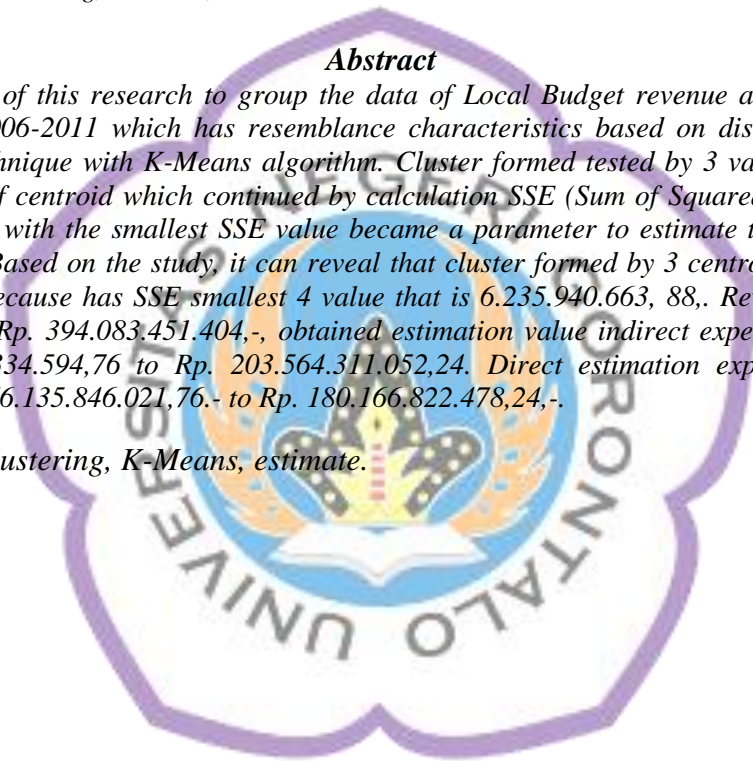
Penelitian ini bertujuan untuk mengelompokkan data APBD tahun 2006-2011 yang memiliki kemiripan karakteristik berdasarkan kedekatan jarak, menggunakan teknik *clustering* dengan algoritma *K-Means*. Pembentukan *cluster* diuji dengan 3 nilai *centroid* dan 2 nilai *centroid*, yang dilanjutkan dengan menghitung nilai *SSE* (Sum of Squared Error). Hasil *cluster* dengan nilai *SSE* terkecil dijadikan sebagai parameter untuk mengestimasi data anggaran belanja yang akan datang. Berdasarkan penelitian yang dilakukan dapat diketahui bahwa pembentukan *cluster* dengan 3 nilai *centroid* adalah *cluster* yang terbaik, karena memiliki nilai *SSE* terkecil yaitu 6.235.940.663,88,. Untuk nilai pendapatan tahun 2012 sebesar Rp 394.083.451.404,-, diperoleh hasil *estimasi* nilai belanja tidak langsung sebesar Rp 199.533.334.595,76,- sampai Rp 203.564.311.052,24. Serta *estimasi* nilai belanja langsung sebesar Rp 176.135.846.021,76,- sampai Rp 180.166.822.478,24,-.

Kata Kunci : *Clustering, K-Means, estimasi.*

Abstract

The objective of this research to group the data of Local Budget revenue and expenditure fiscal year 2006-2011 which has resemblance characteristics based on distance by using clustering technique with K-Means algorithm. Cluster formed tested by 3 value of centroid and 2 value of centroid which continued by calculation SSE (Sum of Squared Error) value. Cluster result with the smallest SSE value became a parameter to estimate the next budget expenditure. Based on the study, it can reveal that cluster formed by 3 centroid value is the best cluster because has SSE smallest 4 value that is 6.235.940.663, 88,. Revenue value on 2012 around Rp. 394.083.451.404,-, obtained estimation value indirect expenditure around Rp. 199.533.334.594,76 to Rp. 203.564.311.052,24. Direct estimation expenditure value around Rp. 176.135.846.021,76.- to Rp. 180.166.822.478,24,-.

Keywords: *clustering, K-Means, estimate.*



DAFTAR ISI

	Halaman
HALAMAN SAMPUL	i
HALAMAN JUDUL	ii
LEMBAR PERNYATAAN	iii
LEMBAR PERSETUJUAN	iv
LEMBAR PENGESAHAN	v
MOTTO DAN PERSEMBAHAN	vi
KATA PENGANTAR	vii
INTISARI	ix
ABSTRACT	x
DAFTAR ISI	xi
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
DAFTAR LAMPIRAN	xv
BAB I PENDAHULUAN	1
A. Latar Belakang.....	1
B. Rumusan Masalah.....	2
C. Ruang Lingkup Penelitian.....	2
D. Tujuan Penelitian.....	3
E. Manfaat penelitian.....	3
BAB II TINJAUAN PUSTAKA	5
A. <i>Knowledge Discovery in Database (KDD)</i> dan <i>Data Mining</i>	5
1. Pengertian data <i>mining</i>	7
2. Tujuan data <i>mining</i>	7
3. Pengelompokkan data <i>mining</i>	8
B. <i>Clustering</i>	11
1. Pengertian <i>clustering</i>	11
2. Metode <i>clustering</i>	12
3. <i>Document clustering</i>	14
4. Klasifikasi algoritma <i>clustering</i>	14
C. Penelitian Terkait.....	15
D. Algoritma <i>K-Means</i>	17
1. Pengertian <i>K-Means</i>	17
2. Algoritma <i>K-Means</i>	18
3. Tahapan algoritma <i>K-Means</i>	20
BAB III METODOLOGI PENELITIAN	23
A. Objek Penelitian.....	23

1. Gambaran umum badan pengelola keuangan dan aset daerah..	23
2. Visi badan pengelola keuangan dan aset daerah	23
3. Misi badan pengelola keuangan dan aset daerah	24
4. Struktur organisasi badan pengelola keuangan dan aset daerah	25
B. Metode Penelitian	26
C. Teknik Pengumpulan Data.....	26
1. Studi literatur.....	26
2. Teknik pengumpulan data arsip	26
D. Tahapan Penelitian.....	27
1. Studi pustaka	28
2. Pengumpulan data arsip	28
3. Praproses data.....	28
4. <i>Clustering</i> menggunakan algoritma <i>K-Means</i>	29
5. Analisis hasil <i>clusterisasi</i>	30
6. Selesai	30
E. Jadwal Penelitian	31
BAB IV HASIL DAN PEMBAHASAN	32
A. Tahapan Penelitian.....	32
1. Pra proses data.....	32
2. <i>Cleaning data</i>	33
3. <i>Clustering</i> menggunakan algoritma <i>K-Means</i>	34
B. Tahap Analisis	35
1. Proses <i>clustering</i> dalam <i>K-Means</i>	35
2. Pola hasil <i>clustering</i> dengan <i>K-Means</i>	57
3. <i>Flowchart clustering K-Means</i>	62
C. Desain Hasil	63
1. Hasil implementasi input data parameter <i>k</i> dan <i>x</i>	63
2. Hasil implementasi input data nama <i>k</i>	64
3. Hasil implementasi input data APBD	65
4. Hasil implementasi tampilan proses hasil perhitungan <i>centroid</i>	66
5. Hasil implementasi perhitungan jarak data ke <i>centroid</i>	67
6. Hasil implementasi <i>iterasi</i>	68
7. Hasil implementasi <i>SSE</i>	69
8. Hasil implementasi laporan data <i>mining</i>	70
9. Hasil implementasi nilai <i>estimasi</i>	71

BAB V PENUTUP	72
A. Kesimpulan	72
B. Saran	72
DAFTAR PUSTAKA	73
LAMPIRAN	



DAFTAR TABEL

Tabel 4.1 Contoh data yang <i>redundant</i>	33
Tabel 4.2 Data APBD Kabupaten XYZ 6 tahun terakhir.....	35
Tabel 4.3 Hasil perhitungan <i>centroid</i> setiap <i>cluster</i> pada pengujian 2 parameter	38
Tabel 4.4 Hasil perhitungan jarak data pada nilai <i>k</i> dengan masing-masing <i>centroid</i> setiap <i>cluster</i>	44
Tabel 4.5 Perhitungan <i>iterasi</i> pertama pada <i>cluster</i> pendapatan	44
Tabel 4.6 Perhitungan <i>iterasi</i> pertama pada <i>cluster</i> belanja tidak langsung.....	45
Tabel 4.7 Perhitungan <i>iterasi</i> pertama pada <i>cluster</i> belanja langsung.....	46
Tabel 4.8 Perhitungan <i>iterasi</i> pertama pada <i>cluster</i> pendapatan.....	48
Tabel 4.9 Perhitungan <i>iterasi</i> pertama pada <i>cluster</i> belanja tidak langsung.....	49
Tabel 4.10 Perhitungan <i>iterasi</i> pertama pada <i>cluster</i> belanja langsung.....	50
Tabel 4.11 Hasil perhitungan jarak data terhadap masing-masing nilai <i>centroid</i> untuk dijadikan seagai perhitungan	53
Tabel 4.12 Perhitungan <i>iterasi</i> kedua pada <i>cluster</i> pendapatan.....	53
Tabel 4.13 Perhitungan <i>iterasi</i> kedua pada <i>cluster</i> belanja tidak langsung	54
Tabel 4.14 Perhitungan <i>iterasi</i> kedua pada <i>cluster</i> belanja langsung	55
Tabel 4.15 Hasil akhir perhitungan <i>iterasi</i> dengan 3 nilai <i>centroid</i> dan 2 nilai <i>centroid</i>	56
Tabel 4.16 Hasil perhitungan nilai <i>SSE</i> dengan 3 nilai <i>centroid</i> dan 2 nilai <i>centroid</i>	57
Tabel 4.17 Data APBD sebelum di <i>cluster</i>	58
Tabel 4.18 Nilai anggota pendapatan, belanja tidak langsung dan belanja langsung pada (C_0)	58
Tabel 4.19 Nilai anggota pendapatan, belanja tidak langsung dan belanja langsung pada (C_1)	58
Tabel 4.20 Nilai anggota pendapatan, belanja tidak langsung dan belanja langsung pada (C_2)	59
Tabel 4.21 Nilai anggota pendapatan 2012.....	60
Tabel 4.22 Contoh nilai <i>outlier</i>	60
Tabel 4.23 Contoh <i>estimasi</i> nilai <i>outlier</i>	61

DAFTAR GAMBAR

Gambar 2.1 Tahapan dalam KDD	5
Gambar 2.2 Contoh <i>clustering</i>	12
Gambar 2.3 Cara kerja algoritma <i>K-Means</i>	21
Gambar 3.1 Tahapan penelitian	27
Gambar 4.1 Flow chart <i>K-Means</i>	62
Gambar 4.2 Implementasi input data parameter <i>k</i> dan <i>x</i> dengan 3 nilai <i>centroid</i>	63
Gambar 4.3 Implementasi input data parameter <i>k</i> dan <i>x</i> dengan 2 nilai <i>centroid</i>	63
Gambar 4.4 Implementasi input data nama <i>k</i> dengan 3 nilai <i>centroid</i>	64
Gambar 4.5 Implementasi input data nama <i>k</i> dengan 2 nilai <i>centroid</i>	64
Gambar 4.6 Implementasi input data APBD dengan 3 nilai <i>centroid</i>	65
Gambar 4.7 Implementasi input data APBD dengan 2 nilai <i>centroid</i>	65
Gambar 4.8 Implementasi tampilan proses hasil perhitungan dengan 3 nilai <i>centroid</i>	66
Gambar 4.9 Implementasi tampilan proses hasil perhitungan dengan 2 nilai <i>centroid</i>	66
Gambar 4.10 Implementasi tampilan perhitungan jarak data dengan 3 nilai <i>centroid</i>	67
Gambar 4.11 Implementasi tampilan perhitungan jarak data dengan 2 nilai <i>centroid</i>	67
Gambar 4.12 Implementasi tampilan <i>iterasi</i> dengan 3 nilai <i>centroid</i>	68
Gambar 4.13 Implementasi tampilan <i>iterasi</i> dengan 2 nilai <i>centroid</i>	68
Gambar 4.14 Implementasi tampilan akhir <i>iterasi</i> dengan 2 nilai <i>centroid</i>	69
Gambar 4.15 Implementasi tampilan nilai <i>SSE</i> dengan 3 nilai <i>centroid</i>	69
Gambar 4.16 Implementasi tampilan nilai <i>SSE</i> dengan 2 nilai <i>centroid</i>	69
Gambar 4.17 Implementasi laporan data <i>mining</i> dengan 3 nilai <i>centroid</i>	70
Gambar 4.18 Implementasi laporan data <i>mining</i> dengan 2 nilai <i>centroid</i>	70
Gambar 4.19 Implementasi tampilan nilai <i>estimasi</i>	71

