

**PENGEMBANGAN ALGORITMA KLASTER DINAMIS
PADA K-MEANS DALAM PENGELOMPOKAN
KINERJA AKADEMIK MAHASISWA
(Studi Kasus: Universitas Pendidikan Ganesha)**

TESIS



PROGRAM STUDI ILMU KOMPUTER
PROGRAM PASCASARJANA
UNIVERSITAS PENDIDIKAN GANESHA
2019

Tesis oleh Komang Ariasa ini telah diperiksa dan disetujui untuk mengikuti Ujian Tesis.


Singaraja, 19 November 2019

Pembimbing I



Dr. I Gede Aris Gunadi, S.Si.,M.Kom.
NIP. 19770318 200812 1 004

Pembimbing II




Prof. Dr. I Made Candiasa, M.I.Kom.
NIP. 19601231 198601 1 004




Tesis oleh Komang Ariasa ini telah berhasil dipertahankan di depan tim penguji dan dinyatakan diterima sebagai sebagian persyaratan untuk memperoleh gelar Magister Komputer di Program Studi Ilmu Komputer, Program Pascasarjana, Universitas Pendidikan Ganesha.


Disetujui pada tanggal : 19 November 2019


oleh Tim Penguji


..... Ketua Dr. Gede Indrawan, ST.,MT.
NIP. 19760102 200312 1 001


..... Anggota Dr. Dewa Gede Hendra Divayana, S.Kom.,M.Kom.
NIP. 19840724 201504 1 002


..... Anggota Dr. Gede Rasben Dantes, S.T.,M.T.I.
NIP. 19750221 200312 1 001



..... Anggota Prof. Dr. I Made Candiasa, M.I.Kom.
NIP. 19601231 198601 1 004


..... Anggota Dr. I Gede Aris Gunadi, S.Si.,M.Kom.
NIP. 19770318 200812 1 004

Mengetahui Direktur

Program Pascasarjana Undiksha,




Prof. Dr. I Gusti Putu Suharta, M.Si.
NIP. 19621215 198803 1 002

LEMBAR PERNYATAAN

Saya menyatakan dengan sesungguhnya, bahwa tesis yang saya susun sebagai syarat untuk memperoleh gelar Magister Ilmu Komputer dari Program Pascasarjana Universitas Pendidikan Ganesha seluruhnya merupakan hasil karya saya sendiri. Adapun bagian-bagian tertentu dalam penulisan tesis yang saya kutip dari hasil karya orang lain telah dituliskan sumbernya secara jelas dan sesuai dengan norma, kaidah, serta etika akademis.

Apabila di kemudian hari ditemukan seluruh atau sebagian tesis ini bukan hasil karya saya sendiri atau adanya plagiat dalam bagian-bagian tertentu, saya bersedia menerima sanksi pencabutan gelar akademik yang saya sandang dan sanksi-sanksi lainnya sesuai dengan peraturan perundang-undangan yang berlaku di wilayah Negara Kesatuan Republik Indonesia.

Singaraja, 19 November 2019
Yang membuat pernyataan,



Komang Ariasa

PRAKATA

Puji dan syukur penulis panjatkan ke hadapan Ida Sang Hyang Widhi Wasa karena atas asung waranugraha-Nya, tesis yang berjudul “**Pengembangan Algoritma Klaster Dinamis Pada K-Means Dalam Pengelompokan Kinerja Akademik Mahasiswa (Studi Kasus: Universitas Pendidikan Ganesha)**” dapat diselesaikan sesuai dengan yang direncanakan.

Tesis ini ditulis untuk memenuhi sebagian persyaratan dalam menyelesaikan studi di Program Pascasarjana Universitas Pendidikan Ganesha, Program Studi Ilmu Komputer. Pada lembar-lembar awal tesis ini, ijin penulis menyampaikan terima kasih dan penghargaan yang sebesar-besarnya kepada pihak-pihak sebagai berikut.

1. Dr. I Gede Aris Gunadi, S.Si.,M.Kom., sebagai pembimbing I yang dengan sabar membimbing, mengarahkan, dan memberikan motivasi, sehingga penulis mampu melewati berbagai kerikil dalam perjalanan studi dan penyelesaian tesis ini.
2. Prof. Dr. I Made Candiasa, M.I.Kom., sebagai pembimbing II yang telah membimbing, mengarahkan, dan memberikan motivasi selama penyusunan tesis, sehingga tesis ini dapat terwujud dengan baik sesuai harapan.
3. Rektor Universitas Pendidikan Ganesha, yang telah memberikan bantuan secara moril dan memfasilitasi berbagai kepentingan studi selama penulis menempuh perkuliahan di Program Pascasarjana Undiksha.
4. Direktur Program Pascasarjana Undiksha dan staf, yang telah banyak membantu selama penulis mengikuti dan menyelesaikan penulisan tesis ini.
5. Ketua program studi dan staf dosen pengajar di program studi Ilmu Komputer yang telah banyak membantu dan memotivasi penulis selama perjalanan studi dan penyusunan tesis ini.
6. Rekan-rekan seangkatan di Program Studi Ilmu Komputer yang dengan karakternya masing-masing telah banyak berkontribusi membentuk kemandirian penulis selama menjalani studi dan menyelesaikan tesis ini.
7. Putu Ardika,S.Pd.,SD dan Luh Nariasih selaku orang tua penulis, serta kakak-kakak, dan keluarga yang selalu memotivasi, banyak membantu secara

material dan moral selama perjalanan studi yang penulis lakoni di Program Studi Ilmu Komputer.

Semoga semua bantuan yang telah mereka taburkan dalam perjalanan studi penulis, mendapat yang sepatasnya oleh Tuhan Yang Maha Esa, sehingga mereka diberi jalan, rezeki, dan keharmonisan dalam menjalani setiap langkah kehidupan.

Penulis menyadari sepenuhnya bahwa laporan tesis ini masih jauh dari sempurna, untuk itu semua jenis saran, kritik dan masukan yang bersifat membangun sangat penulis harapkan. Akhir kata, semoga tulisan ini dapat memberikan manfaat dan memberikan wawasan tambahan bagi para pembaca.

Singaraja, 19 November 2019



Penulis

DAFTAR ISI

	HALAMAN
PRAKATA	vi
ABSTRAK	viii
DAFTAR ISI	x
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiv
DAFTAR LAMPIRAN	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang Penelitian	1
1.2 Identifikasi Masalah	6
1.3 Pembatasan Masalah	6
1.4 Rumusan Masalah	7
1.5 Tujuan Penelitian.....	7
1.6 Manfaat Penelitian.....	8
BAB II KAJIAN TEORI	9
2.1 <i>Data Mining</i>	9
2.2 CRISP-DM	12
2.3 <i>Clustering</i>	14
2.4 Algoritma <i>K-Means</i>	15
2.5 <i>Dynamic K-Means</i>	17
2.6 Ukuran Validitas <i>Cluster</i>	19
2.7 Penelitian Terkait	25
2.8 Konsep Penelitian Menggunakan Metode <i>K-Means</i> Dinamis Dengan Inisialisasi <i>Centroid</i>	30
2.8.1 <i>Initialize Cluster</i>	30
2.8.2 <i>K-Means Algorithm</i>	31
2.8.3 <i>Cluster Variance</i>	32

BAB III METODE PENELITIAN	33
3.1 Rencana Alur Penelitian.....	33
3.2 Prosedur Pengumpulan Data	35
3.3 Rancangan Penelitian	37
3.4 Teknik Analisis Data	44
BAB IV HASIL PENELITIAN DAN PEMBAHASAN.....	45
4.1 Pembentukan Model dan Karakteristik Data Penelitian.....	45
4.1.1 Pembersihan Data	45
4.1.2 Integrasi Data	47
4.1.3 Reduksi Data.....	47
4.2 Implementasi K-Means Dinamis.....	52
4.3 Hasil Evaluasi Klustering	57
4.3.1 Pengujian Pertama	58
4.3.2 Pengujian Kedua.....	65
4.3.3 Pengujian Ketiga.....	71
4.3.4 Uji Perbandingan Hasil Evaluasi Klustering	77
4.4 Analisis Persebaran Data Klustering	87
4.4.1 Klaster 1.....	89
4.4.2 Klaster 2.....	91
4.4.3 Klaster 3.....	93
4.4.4 Klaster 4.....	95
4.4.5 Klaster 5.....	97
BAB V PENUTUP	99
5.1 Rangkuman.....	99
5.2 Simpulan.....	102
5.3 Saran.....	104
DAFTAR RUJUKAN.....	106
LAMPIRAN.....	109

DAFTAR TABEL

Tabel	Halaman
Tabel 3.1 Karakteristik Data Penelitian	40
Tabel 4.1 Profil Data Berdasarkan Kriteria Jenis Kelamin.....	47
Tabel 4.2 Profil Data Kriteria Usia	48
Tabel 4.3 Profil Data Kriteria Alamat Asal	48
Tabel 4.4 Hasil Inisiasi Kriteria Alamat Asal	50
Tabel 4.5 Hasil Inisiasi Kriteria Status Sekolah.....	51
Tabel 4.6 Hasil Inisiasi Kriteria Jalur Masuk Mahasiswa.....	52
Tabel 4.7 Data Kriteria dan Jumlah Pekerjaan Orang Tua Mahasiswa	52
Tabel 4.8 Algoritma Inisiasi K-Means Dengan Metode Mean	54
Tabel 4.9 Algoritma Pengujian Cluster Variance	56
Tabel 4.10 Inisiasi Centroid Awal Pengujian Pertama Dengan Jumlah 2 Klaster	58
Tabel 4.11 Inisiasi Centroid Awal Pengujian Pertama Dengan Jumlah 3 Klaster	58
Tabel 4.12 Inisiasi Centroid Awal Pengujian Pertama Dengan Jumlah 4 Klaster	58
Tabel 4.13 Inisiasi Centroid Awal Pengujian Pertama Dengan Jumlah 5 Klaster	59
Tabel 4.14 Hasil Pengujian Pertama Evaluasi Algoritma Dengan 765 Data.....	59
Tabel 4.15 Hasil SSE dan Jumlah Iterasi Pengujian Pertama.....	61
Tabel 4.16 Hasil DBI dan Jumlah Iterasi Pengujian Pertama.....	62
Tabel 4.17 Hasil PC dan Jumlah Iterasi Pengujian Pertama.....	63
Tabel 4.18 Hasil Variance dan Jumlah iterasi Pengujian Pertama.....	64
Tabel 4.19 Inisiasi Centroid Awal Pengujian Kedua Dengan Jumlah 2 Klaster...	65
Tabel 4.20 Inisiasi Centroid Awal Pengujian Kedua Dengan Jumlah 3 Klaster...	65
Tabel 4.21 Inisiasi Centroid Awal Pengujian Kedua Dengan Jumlah 4 Klaster...	65
Tabel 4.22 Inisiasi Centroid Awal Pengujian Kedua Dengan Jumlah 5 Klaster...	65
Tabel 4.23 Hasil Pengujian Kedua Evaluasi Algoritma Dengan 765 Data.....	66
Tabel 4.24 Hasil SSE dan Jumlah Iterasi Pengujian Kedua.....	67
Tabel 4.25 Hasil DBI dan Jumlah Iterasi Pengujian Kedua.....	68
Tabel 4.26 Hasil PC dan Jumlah Iterasi Pengujian Kedua.....	69
Tabel 4.27 Hasil Variance dan Jumlah Iterasi Pengujian Kedua	70
Tabel 4.28 Inisiasi Centroid Awal Pengujian Ketiga Dengan Jumlah 2 Klaster ..	71

Tabel 4.29 Inisiasi Centroid Awal Pengujian Ketiga Dengan Jumlah 3 Klaster ..	71
Tabel 4.30 Inisiasi Centroid Awal Pengujian Ketiga Dengan Jumlah 4 Klaster ..	71
Tabel 4.31 Inisiasi Centroid Awal Pengujian Ketiga Dengan Jumlah 5 Klaster ..	72
Tabel 4.32 Hasil Pengujian Ketiga Evaluasi Algoritma Dengan 765 Data	72
Tabel 4.33 Hasil SSE dan Jumlah Iterasi Pengujian Ketiga	74
Tabel 4.34 Hasil PC dan Jumlah Iterasi Pengujian Ketiga	75
Tabel 4.35 Hasil PC dan Jumlah Iterasi Pengujian Ketiga	76
Tabel 4.36 Hasil Variance dan Jumlah Iterasi Pengujian Ketiga.....	77
Tabel 4.37 Nilai Centroid Awal K-means Dinamis Dengan Mean Based.....	82
Tabel 4.38 Nilai Validasi K-Means Klaster Dinamis Dengan Inisiasi Klaster.....	83
Tabel 4.39 Nilai Titik Pusat Inisiasi K-Means Dinamis Dengan Mean Based.....	84
Tabel 4.40 Perbandingan K-Means Dinamis Mean Based Dengan Algoritma Lain	85



DAFTAR GAMBAR

Gambar	Halaman
Gambar 2.1 Data Mining Sebagai Salah Satu Tahap Dalam Proses Discovery Knowledge	11
Gambar 2.2 Siklus Hidup CRISP-DM	12
Gambar 2.3 Cabang Teknik Clustering.....	14
Gambar 2.4 Flowcart Algoritma Dynamic K-Means.....	17
Gambar 2.5 Algoritma K-Means Kluster Dinamis dengan Inisialisasi Kluster	30
Gambar 3.1 Rencana Alur Penelitian.....	34
Gambar 3.2 Algoritma Kluster Dinamis Pada K-Means	42
Gambar 4.1 Perbandingan Nilai SSE Pengujian Pertama.....	61
Gambar 4.2 Grafik Nilai Davies Bouldin Index Pengujian Pertama	63
Gambar 4.3 Grafik Perbandingan Nilai Purity Coefficient Pengujian Pertama	64
Gambar 4.4 Perbandingan Nilai SSE Pengujian Kedua	68
Gambar 4.5 Grafik Perbandingan Nilai Davies Bouldin Index Pengujian Kedua	69
Gambar 4.6 Grafik Perbandingan Nilai Purity Coefficient Pengujian Kedua.....	70
Gambar 4.7 Perbandingan Nilai SSE Pengujian Ketiga	74
Gambar 4.8 Grafik Perbandingan Nilai Davies Bouldin Index Pengujian Ketiga	75
Gambar 4.9 Grafik Perbandingan Nilai Purity Coefficient Pengujian Ketiga	77
Gambar 4.10 Perbandingan Rata-Rata Nilai DBI Ketiga Algoritma.....	78
Gambar 4.11 Perbandingan Rata-Rata Nilai SSE Ketiga Algoritma.....	79
Gambar 4.12 Perbandingan Rata-Rata Nilai PC Ketiga Algoritma.....	80
Gambar 4.13 Perbandingan Rata-Rata Nilai Variance Kluster Dinamis	82
Gambar 4.14 Nilai Cluster Variance K-Means Dinamis Dengan Mean Based	83
Gambar 4.15 Nilai Validitas Jumlah Kluster K-Means Dinamis Mean Based	84
Gambar 4.16 Hasil Klastering Dengan K-Means Dinamis.....	88
Gambar 4.17 Hasil Persebaran Kluster 1	90
Gambar 4.18 Hasil Persebaran Kluster 2	92
Gambar 4.19 Hasil Persebaran Kluster 3	94
Gambar 4.20 Hasil Persebaran Kluster 4	96
Gambar 4.21 Hasil Persebaran Kluster 5	98

DAFTAR LAMPIRAN

Lampiran	Halaman
Lampiran 1 Contoh Proses Perhitungan K-Means Dengan Inisialisasi Centroid	110
Lampiran 2 Perbandingan Hasil Pengujian Dengan Nilai Sum of Square Error Minimum (SSE).....	117
Lampiran 3 Perbandingan Hasil Pengujian Dengan Nilai <i>Davies Bouldin Index (DBI)</i>	118
Lampiran 4 Perbandingan Hasil Pengujian Dengan Nilai Purity Coeficient (PC)	119
Lampiran 5 Perbandingan Hasil Pengujian Dengan Nilai Variance.....	120
Lampiran 6 Source Code Implementasi Algoritma	121
Lampiran 7 Data Awal Perhitungan Klastering.....	124

