

SKRIPSI
PENERAPN METODE *K-MEANS* UNTUK
PENGELOMPOKAN JUMLAH DATA AKTA KELAHIRAN
BERDASARKAN PROVINSI DI INDOESIA



Oleh

Gloriana Ernawati Bulu

19083000168

PROGRAM S1 SISTEM INFORMASI
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS MERDEKA
MALANG
2023

**PENERAPAN METODE *K-MEANS* UNTUK
PENGELOMPOKAN JUMLAH DATA AKTA KELAHIRAN
BERDASARKAN PROVINSI DI INDONESIA**



SKRIPSI

Diajukan Sebagai Salah Satu Syarat
Memperoleh Gelar Sarjana System Informasi
Pada Program S1 Sistem Informasi
Universitas Merdeka Malang

Oleh

Gloriana Ernawati Bulu

19083000168

**PROGRAM S1 SISTEM INFORMASI
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS MERDEKA
MALANG**

2023

LEMBAR PERSETUJUAN

Nama : Gloriana Ernawati Bulu
NIM : 19083000168
Universitas : Universitas Merdeka Malang
Fakultas : Teknologi Informasi
Program Studi : S1 Sistem Informasi
Judul : Penerapan Metode *K-Means* Untuk
Pengelompokan Jumlah Data Akta Kelahiran
Berdasarkan Provinsi Di Indonesia

Malang, 24 Juli 2023

Disetujui Dan Diterima

Ketua Program Studi S1 Sistem Informasi



Galandaru Swafaganata, S.Si., M.Si
NIDN. 728109104

Dosen Pembimbing



Drs. Anis Zubair, M.Kom.
NIDN. 701116703

Dekan
Fakultas Teknologi Informasi

M. Martini Andarwati, SE., M.Si
NIDN. 7037601

**PENERAPAN METODE *K-MEANS* UNTUK
PENGELOMPOKAN JUMLAH DATA AKTA KELAHIRAN
BERDASARKAN PROVINSI DI INDONESIA**

Dipersiapkan dan disusun oleh
Gloriana Ernawati Bulu
19083000168

Telah dipertahankan di depan Dewan Penguji
Pada Tanggal 24 Juli 2023

Susunan Dewan Penguji

Ketua Penguji



Himawan Pramaditya, S.Kom., M.Kom
NIDN. 716098405

Sekretaris Penguji



Drs. Agis Zubair, M.Kom
NIDN. 701116703

Anggota Penguji



Viry Puspaning Ramadhan, A.Nd.Li., S.Kom., M.Kom
NIDN. 725019602

Skripsi Ini Telah Diterima Sebagai Salah Satu Persyaratan untuk
Memperoleh Gelar Sarjana Sisten Informasi

Malang, 24 Juli 2023

Dekan Fakultas Teknologi Informasi
Universitas BKRDEKA Malang



Dr. Matriana Andarwati, SE., M.Kom
NIDN. 716037601

SURAT PERNYATAAN

Yang bertanda tangan dibawah ini

Nama : Gloriana Ernawati Bulu
Nim : 19083000168
Program Studi : SI Sistem Informasi
Bidang Kajian Skripsi : Analisis Data
Judul Tugas Akhir : Penerapan Metode *K-Means* Untuk Pengelompokan Jumlah Data Akta Kelahiran Berdasarkan Provinsi Di Indonesia
Lokasi Penelitian : Badan Pusat Statistik (BPS) Indonesia
Alamat Rumah Asal : Jl. Simpang Mega Mendung No 50
No. Telp/Email : 082139715636
wulanlamend@gmail.com

Dengan ini menyatakan bahwa saya benar-benar melakukan penelitian dan penulisan skripsi tersebut diatas benar-benar karya saya dan tidak melakukan plagiasi. Kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam dengan disebutkan nama dan dicantumkan dalam daftar pustaka.

Demikian surat pernyataan ini saya buat dengan sesungguhnya dan apabila dikemudian hari terdapat penyimpangan dan tidak kebenaran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik sesuai dengan aturan yang berlaku.

Malang, 24 Juli 2023



Gloriana Ernawati Bulu

MOTO

“ TIDAK ADA MIMPI YANG TERLALU TINGGI, TIDAK ADA MIMPI
YANG PATUT UNTUK DIREMEHKAN, LAMBUNGAN SETINGGI YANG
KAU INGINKAN DAN GAPAILAH DENGAN SELAYAKNYA YANG KAU
HARAPKAN “

LEMBAR PERSEMBAHAN

Skripsi ini penulis persembahkan untuk kedua orang tua, kakak, adik dan keluarga yang sudah memberikan dukungan penuh bagi penulis dari masih kecil hingga sekarang. Yang sudah dengan susah payah dalam pengorbanan, doa, usaha hingga cinta dan kasih sayang yang telah diberikan. Serta tak lupa untuk semua sahabat dekat penulis yang ikut membantu dalam memberi dukungan kepada penulis. Terima kasih banyak penulis ucapkan kepada seluruh orang-orang yang telah memberikan dukungan penuh.

KATA PENGANTAR

Puji dan Syukur penulis haturkan kehadirat Tuhan Yang Maha Esa, karena berkat Karunia-Nya, tugas akhir ini dapat diselesaikan. Pada kesempatan yang baik ini dan dengan penuh rasa syukur yang mendalam, perkenankan penulis mengucapkan terimakasih yang sebesar-besarnya dan apresiasi yang setinggi-tingginya kepada yang terhormat:

1. Ibu Dr. Mardiana Andarwati, SE., M.Si selaku Ketua Dekan Program Studi Sistem Informasi yang telah memberikan dedikasi selama 4 tahun untuk tercapainya pembelajaran maksimal bagi program studi yang dipimpin.
2. Bapak Galandaru Swalaganata, S.Si., M.Si selaku Ketua Program Studi S1 Sistem Informasi Fakultas Teknologi Informasi Universitas Merdeka Malang.
3. Bapak Drs, Anis Zubair, M.Kom selaku dosen pembimbing saya yang telah memberikan arahan, bimbingan, serta motivasi kepada penulis dan terima kasih atas kesabaran beliau kepada penyusun selama proses pengerjaan skripsi ini.
4. Yang paling istimewa yaitu kepada kedua orang tua saya yang menjadi kekuatan terbesar bagi penulis, terima kasih telah memberikan segenap cinta dan kasih sayang, doa yang tidak pernah putus, motivasi dan memberikan dukungan penuh hingga penulis bisa menyelesaikan skripsi ini.
5. Kiyut Family yang telah membantu dan memberi semangat selama proses pengerjaan skripsi ini.
6. Teman kelas dan saudara di tanah rantauan Elen, Vhyn, yang sudah berjuang bersama dengan caranya masing-masing membantu penulis selama menyusun skripsi ini.
7. Frengky Ata Goran yang telah memberikan motivasi dan mensupport penulis sehingga skripsi ini dapat terselesaikan.
8. Untuk ketujuh member BTS, Kim Namjoon, Kim Seokjin, Min Yoongi, Jung Hoseok, Park Jimin, Kim Taehyung, Jeon Jungkook yang secara tidak langsung menjadi penyemangat penulis dalam menyelesaikan skripsi ini.

DAFTAR ISI

HALAMAN SAMPUL.....	i
HALAMAN JUDUL	ii
LEMBAR PERSETUJUAN.....	iii
LEMBAR PENGESAHAN	iv
SURAT PERNYATAAN	v
MOTO.....	vi
LEMBAR PERSEMBAHAN	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR.....	xii
DAFTAR LAMPIRAN	xiii
ABSTRAK	xiv
ABCTRAC	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Tujuan Penelitian	2
1.4 Manfaat Penelitian	3
1.5 Batasan Penelitian.....	3
BAB II KAJIAN PUSTAKA	4
2.1 Landasan Teori	4
2.1.1 <i>Data Mining</i>	4
2.1.2 <i>Orange</i>	6
2.1.3 <i>K-Means</i>	6
2.1.4 <i>Clustering</i>	6
2.1.5 Akta Kelahiran.....	8
2.1.6 <i>Silhouette Score</i>	8
2.2 Penelitian Terdahulu.....	9

BAB III METODE PENELITIAN	11
3.1 Desain Penelitian	11
3.2 Lokasi Penelitian	12
3.3 Populasi dan Sampel atau Objek Penelitian	12
3.3.1 Populasi	12
3.3.2 Sampel	12
3.4 Teknik Pengumpulan Data	13
3.4.1 BPS	13
3.4.2 Studi Pustaka	13
3.5 Teknik Analisis Data	13
BAB IV HASIL DAN PEMBAHASAN	11
4.1 Hasil Penelitian	14
4.2 Pembahasan	16
4.2.1 Perhitungan <i>Clustering</i> secara manual	16
4.2.2 Perhitungan <i>Clustering</i> Menggunakan <i>Orange</i>	32
BAB V PENUTUP	43
5.1 Kesimpulan	43
5.2 Saran	43
DAFTAR PUSTAKA	44
DAFTAR LAMPIRAN	45

DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu	9
Tabel 4. 1 Data Awal Presentase Akta Kelahiran	14
Tabel 4. 2 Proporsi Anak Dibawah 5 Tahun.....	15
Tabel 4. 3 Keeratan Hubungan.....	16
Tabel 4. 4 Hasil Korelasi.....	17
Tabel 4. 5 Penentuan <i>Centroid</i> Awal Tahun 2018.....	18
Tabel 4. 6 Penentuan Jarak Terdekat	18
Tabel 4. 7 Hasil <i>Clustering</i> Manual 2018	19
Tabel 4. 8 Penentuan <i>Centroid</i> Awal Tahun 2019	20
Tabel 4. 9 Perhitungan Jarak Terdekat.....	21
Tabel 4. 10 Hasil <i>Clustering</i> Manual 2019	22
Tabel 4. 11 Penentuan <i>Centroid</i> Awal Tahun 2020.....	23
Tabel 4. 12 Perhitungan Jarak Terdekat.....	24
Tabel 4. 13 Hasil <i>Clustering</i> Manual 2019	25
Tabel 4. 14 Penentuan <i>Centroid</i> Awal Tahun 2021	26
Tabel 4. 15 Perhitungan Jarak Terdekat.....	27
Tabel 4. 16 Hasil <i>Clustering</i> Manual 2021	29
Tabel 4. 17 Penentuan <i>Centroid</i> Awal Tahun 2022.....	30
Tabel 4. 18 Perhitungan Jarak Terdekat.....	31
Tabel 4. 19 Hasil <i>Clustering</i> Manual 2022	32

DAFTAR GAMBAR

Gambar 2. 1 Tahap-Tahap Data <i>Mining</i>	5
Gambar 3. 1 <i>Flowchart</i> proses penelitian	11
Gambar 4. 1 Silhouette Scores Tahun 2018.....	35
Gambar 4. 2 Hasil <i>Clustering</i> Menggunakan <i>Orange</i> 2018	34
Gambar 4. 3 Silhouette Scores Tahun 2019.....	35
Gambar 4. 4 Hasil <i>Clustering</i> Menggunakan <i>Orange</i> 2019	36
Gambar 4. 5 Silhouette Scores Tahun 2020.....	37
Gambar 4. 6 Hasil <i>Clustering</i> Menggunakan <i>Orange</i> 2020	38
Gambar 4. 7 Silhouette Scores Tahun 2021	39
Gambar 4. 8 Hasil <i>Clustering</i> Menggunakan <i>Orange</i> 2021	40
Gambar 4. 9 Silhouette Scores Tahun 2022.....	40
Gambar 4. 10 Hasil <i>Clustering</i> Menggunakan <i>Orange</i> 2022	42

DAFTAR LAMPIRAN

Lampiran 1. Daftar Riwayat Hidup.....	45
Lampiran 4. Hasil Plagiasi	46
Lampiran 5. Lembar Persetujuan Tugas Akhir	47

PENERAPAN METODE *K-MEANS* UNTUK PENGELOMPOKAN JUMLAH DATA AKTA KELAHIRAN BERDASARKAN PROVINSI DI INDONESIA

Gloriana Ernawati Bulu

Program Studi Teknologi Informasi, Program
Sistem Informasi Universitas Merdeka Malang

Email : wulanlamend@gmail.com

ABSTRAK

Akta kelahiran adalah Akta Catatan Sipil hasil pencatatan kelahiran bayi yang dilaporkan kelahirannya akan terdaftar dalam Kartu Keluarga dan diberi Nomor Induk Kependudukan (NIK) sebagai dasar untuk memperoleh pelayanan publik. Salah satu permasalahan yang dihadapi Kementerian Kependudukan dan Pencatatan Sipil (Disdukcapil) adalah kepemilikan akta kelahiran yang tidak sebanding dengan banyaknya jumlah penduduk di Indonesia. Penelitian ini dengan untuk mengelompokkan data akta kelahiran berdasarkan jumlah provinsi yang ada di Indonesia menggunakan metode *K-Means* dan untuk menginterpretasikan hasil *Clustering* data akta kelahiran untuk mengambil informasi atau insight yang berguna. Metode yang digunakan dalam penelitian adalah *K-Means Clustering*, merupakan metode data *Clustering* non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster* atau kelompok. Berdasarkan hasil pada pengelompokan data akta kelahiran di Indonesia dan proporsi anak dibawah 5 tahun, diperoleh hasil pada tahun 2018 diperoleh *cluster* tertinggi (C1) terdiri dari 23 provinsi dan *cluster* terendah (C2) 11 provinsi, di tahun 2019 diperoleh *cluster* tertinggi (C1) terdiri dari 25 provinsi dan *cluster* terendah 9 provinsi, di tahun 2020 diperoleh *cluster* tertinggi (C1) terdiri dari 26 provinsi dan *cluster* terendah 8 provinsi, di tahun 2021 diperoleh *cluster* tertinggi (C1) terdiri dari 30 provinsi dan *cluster* terendah 4 provinsi, dan tahun 2022 diperoleh *cluster* tertinggi (C1) terdiri dari 31 provinsi dan *cluster* terendah 3 provinsi. Hasil ini dapat dijadikan acuan bagi Dinas Kependudukan dan Pencatatan Sipil (Disdukcapil) dalam melakukan pemetaan data akta kelahiran di Indonesia, sehingga kepemilikan akta kelahiran di Indonesia dapat sebanding dengan jumlah penduduk yang ada di di Indonesia.

Kata Kunci : Akta Kelahiran, *K-Means*, *Clustering*

*APPLICATION OF K-MEANS METHOD FOR CLUSTERING THE NUMBER
OF BIRTH CERTIFICATE DATA BASED ON PROVINCES IN INDONESIA*

Gloriana Ernawati Bulu

Program Studi Teknologi Informasi, Program
Sistem Informasi Universitas Merdeka Malang

Email : wulanlamend@gmail.com

ABSTRACT

A birth certificate is a Civil Record Certificate resulting from the recording of the birth of a baby whose birth is reported to be registered in the Family Card and given a Population Identification Number (NIK) as a basis for obtaining public services. One of the problems faced by the Ministry of Population and Civil Registration (Disdukcapil) is the ownership of birth certificates that are not proportional to the large number of people in Indonesia. This research is to cluster birth certificate data based on the number of provinces in Indonesia using the K-Means method and to interpret the results of clustering birth certificate data to retrieve useful information or insight. The method used in the research is K-Means Clustering, which is a non-hierarchical data clustering method that attempts to partition existing data into the form of one or more clusters or groups. Based on the results on the clustering of birth certificate data in Indonesia and the proportion of children under 5 years, the results obtained in 2018 obtained the highest cluster (C1) consisting of 23 provinces and the lowest cluster (C2) 11 provinces, in 2019 obtained the highest cluster (C1) consisting of 25 provinces and the lowest cluster 9 provinces, in 2020 obtained the highest cluster (C1) consisting of 26 provinces and the lowest cluster 8 provinces, in 2021 obtained the highest cluster (C1) consisting of 30 provinces and the lowest cluster 4 provinces, and in 2022 obtained the highest cluster (C1) consisting of 31 provinces and the lowest cluster 3 provinces. These results can be used as a reference for the Population and Civil Registration Office (Disdukcapil) in mapping birth certificate data in Indonesia, so that birth certificate ownership in Indonesia can be proportional to the population in Indonesia.

Keywords: Birth Certificate, K-Means, Clustering