

**KLUSTERISASI SISWA SEKOLAH DASAR DENGAN
ALGORITMA *K-MEANS CLUSTERING***

Skripsi

**Disusun untuk memenuhi salah satu syarat
memperoleh gelar Sarjana Komputer**



**Rachel Haryawan
1313618026**


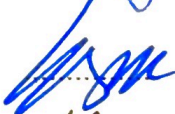





**PROGRAM STUDI ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI JAKARTA
2023**

LEMBAR PERSETUJUAN HASIL SIDANG SKRIPSI

KLUSTERISASI SISWA SEKOLAH DASAR DENGAN ALGORITMA K-MEANS CLUSTERING

Nama : Rachel Haryawan

No. Registrasi : 1313618026

	Nama	Tanda Tangan	Tanggal
Penanggung Jawab			
Dekan	: <u>Prof. Dr. Muktiningsih N, M.Si.</u> NIP. 196405111989032001		28/8-23
Wakil Penanggung Jawab			
Wakil Dekan I	: <u>Dr. Esmar Budi, S.Si., MT.</u> NIP. 197207281999031002		28/8-23
Ketua	: <u>Med Irzal, M.Kom.</u> NIP. 197706152003121001		22/8-23
Sekretaris	: <u>Ari Hendarno, S.Pd, M.Kom</u> NIP. 198811022022031002		22/8-23
Penguji	: <u>Ir. Fariani Hermin Indiyah, MT.</u> NIP. 196002111987032001		22/8-23
Pembimbing I	: <u>Dr. Ria Arafiyah, M.Si.</u> NIP. 197511212005012004		23/8-23
Pembimbing II	: <u>Drs. Mulyono, M.Kom.</u> NIP. 196605171994031003		23/8-23

Dinyatakan lulus ujian skripsi tanggal: 18 Agustus 2023

LEMBAR PERNYATAAN

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul “Klusterisasi Siswa Sekolah Dasar dengan Algoritma *K-Means Clustering*” yang disusun sebagai syarat untuk memperoleh gelar Sarjana Komputer dari Program Studi Ilmu Komputer Universitas Negeri Jakarta adalah karya ilmiah saya dengan arahan dari dosen pembimbing.

Sumber informasi yang diperoleh dari penulis lain yang telah dipublikasikan yang disebutkan dalam teks skripsi ini, telah dicantumkan dalam Daftar Pustaka sesuai dengan norma, kaidah dan etika penulisan ilmiah.

Jika di kemudian hari ditemukan sebagian besar skripsi ini bukan hasil karya saya sendiri dalam bagian-bagian tertentu, saya bersedia menerima sanksi pencabutan gelar akademik yang saya sanding dan sanksi-sanksi lainnya sesuai dengan peraturan perundang-undangan yang berlaku.

Bekasi, 3 Agustus 2023



Rachel Haryawan

HALAMAN PERSEMBAHAN



Untuk Mama, Papa dan Iie Lan, terima kasih atas kasih sayang dan cinta yang terus – menerus diberikan. Skripsi ini ada karena kasih sayang, dukungan dan doa dari kalian.

Dan untuk diriku sendiri, terima kasih karena memilih untuk bertahan.

ABSTRAK

RACHEL HARYAWAN. Clusterisasi Siswa Sekolah Dasar dengan Algoritma *K-Means Clustering*. Skripsi, Program Studi Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Jakarta. Agustus 2023.

Dalam pendidikan formal seperti sekolah, tidak dipungkiri nilai merupakan sesuatu yang penting untuk siswa. Selain itu penilaian juga dapat dijadikan sebagai standar untuk menentukan peringkat siswa di sekolah. Metode pembelajaran yang digunakan di kebanyakan sekolah saat ini merupakan metode *one-size-fit-all*, namun metode ini kurang efektif karena menyamaratakan semua siswa. Penelitian ini dibuat agar dapat mendapatkan kelompok siswa sehingga nantinya dapat digunakan oleh para guru untuk menentukan metode pembelajaran yang terbaik berdasarkan pola masing – masing kluster siswa. Penelitian ini menggunakan *K-Means Clustering* untuk mengelompokkan siswa berdasarkan nilai dan data latar belakang mereka untuk membuat suatu model kluster yang dapat dianalisis setiap klusternya. Jika sudah diketahui kelompok siswa berdasarkan nilai dan latar belakangnya, maka hasilnya dapat digunakan oleh para guru untuk memberikan penanganan yang tepat bagi setiap kelompok siswa tersebut. Data yang digunakan sebanyak 49 kolom dan 377 baris. Sebelum melakukan proses klusterisasi, data melalui tahap *pre – processing* terlebih dahulu. Data yang didapat setelah *pre – processing* sebanyak 37 kolom dan 365 baris. Kemudian, data dibagi ke dalam tiga kluster yang divalidasi oleh *Elbow Method*. Hasil klusterisasi membagi siswa ke dalam kluster 1 sebanyak 132 orang, kluster 2 sebanyak 112 orang dan kluster 3 sebanyak 112 orang. Hasil analisis kluster menunjukkan bahwa kluster 1 lebih unggul dalam rata – rata nilai keterampilan serta KI1 dan KI2.

Kata kunci: Machine Learning, Kurikulum 2013, Algoritma Klusterisasi, Klusterisasi Siswa

ABSTRACT

RACHEL HARYAWAN. Elementary School Clustering with K-Means Clustering Algorithm. Thesis, Computer Science, Faculty of Mathematics and Natural Sciences, State University of Jakarta. August 2023.

In the formal education such as school, we can't deny that a score is still an important thing for students. In addition, assessment can also be used as a standard for determining student rankings in schools. A method that mostly used by school nowadays is a one-size-fit-all method, but this method is not too effective because treats all students same or equally. This research made so that the cluster of students can be found and the teachers can use it to find the effective learning method for every pattern of student clusters. This research is using K-Means Clustering for grouping students based on their scores and background data to create a cluster model that can be analyzed by each cluster. If the group of students is already known based on their scores and background, then the results can be used by teachers to provide appropriate treatment for each group of students. The data used in this research contains 49 columns and 377 rows. Before entering the clustering process, the data goes through the pre-processing stage first. The data after pre – processing contains 37 columns and 365 rows. Then the data was divided into three clusters validated by the Elbow Method. The clustering results divided students into cluster 1 with 132 people, cluster 2 with 112 people and cluster 3 with 121 people. The results of the cluster analysis showed that cluster 1 was superior in the average value of skills and KI1 and KI2.

Keywords: Machine Learning, 2013 Curriculum, Clustering Algorithm, Students Clustering

KATA PENGANTAR

Puji dan syukur peneliti panjatkan kehadirat Tuhan Yang Maha Esa, karena atas rahmat dan karunia-Nya peneliti dapat menyelesaikan skripsi ini dengan baik. Skripsi ini disusun sebagai syarat untuk lulus di Universitas Negeri Jakarta

Peneliti menyadari akan adanya ketidaksempurnaan penulisan baik dari segi materi, segi teknis, maupun segi bahasa yang digunakan. Terwujudnya skripsi ini berkat adanya dukungan, bimbingan dan dorongan dari Bapak/Ibu Dosen, orang tua serta teman – teman yang telah membantu proses penyusunan skripsi ini hingga terselesaikannya dengan baik. Peneliti ingin mengucapkan terima kasih kepada pihak – pihak yang telah membantu terwujudnya skripsi ini:

1. Dr. Ria Arafiyah, M.Si, Koordinator Program Studi Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Jakarta dan selaku dosen pembimbing pertama yang telah memberikan banyak arahan dan bimbingan dalam proses pembuatan skripsi ini sehingga skripsi ini dapat selesai dengan baik.
2. Drs. Mulyono, M.Kom. selaku dosen pembimbing kedua yang telah memberikan banyak arahan dan bimbingan dalam proses pembuatan skripsi ini sehingga proposal skripsi ini dapat selesai dengan baik.
3. Seluruh dosen Ilmu Komputer Universitas Negeri Jakarta yang telah memberikan ilmu kepada peneliti selama peneliti menempuh pendidikan di Universitas Negeri Jakarta.
4. Orang tua peneliti, yang telah memberikan dukungan, doa, kasih sayang serta motivasi selama peneliti mengerjakan skripsi ini sehingga skripsi ini dapat diselesaikan dengan baik

5. Seluruh teman – teman dari program studi Ilmu Komputer Universitas Negeri Jakarta angkatan 2018 atas dukungannya sebagai teman dan kerabat selama peneliti menempuh pendidikan di Universitas Negeri Jakarta.
6. Annisa Salsabilla, Petra Renatta, Kak Fanuel Vanessa, dan Kak Febrirani sebagai sahabat yang memberikan dukungan, motivasi, dan doa selama peneliti mengerjakan skripsi ini.
7. Irvhan Christov, untuk seluruh motivasi dan dukungan yang telah diberikan selama peneliti mengerjakan skripsi ini, baik secara moril maupun material.
8. Diah Anggraeny, Nilatil Moena dan Novitasari sebagai sahabat yang telah banyak membantu dan mendukung peneliti selama berkuliah di Universitas Negeri Jakarta dan selama mengerjakan skripsi ini.

Harapan peneliti, semoga skripsi ini dapat bermanfaat bagi pembaca dan bagi pihak – pihak terkait. Peneliti mengharapkan kritik dan saran sehingga skripsi ini bisa lebih baik dan bermanfaat untuk masa – masa yang akan datang.

Jakarta, 2023

Peneliti

DAFTAR ISI

LEMBAR PERSETUJUAN HASIL SIDANG SKRIPSI.....	ii
LEMBAR PERNYATAAN.....	iii
HALAMAN PERSEMBAHAN.....	iv
ABSTRAK.....	v
ABSTRACT.....	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiii
DAFTAR LAMPIRAN.....	xv
BAB I PENDAHULUAN.....	1
A. Latar Belakang.....	1
B. Perumusan Masalah.....	4
C. Tujuan Penelitian.....	4
D. Batasan Masalah.....	4
E. Manfaat Penelitian.....	5
BAB II KAJIAN PUSTAKA.....	5
A. <i>Machine Learning</i>.....	5
B. <i>Clustering dan Algoritma K-Means Clustering</i>.....	6
C. <i>Elbow Method</i>.....	8
D. <i>Principal Component Analysis (PCA)</i>.....	9
E. Kurikulum 2013.....	10
1. Penilaian Sikap.....	10
2. Penilaian Pengetahuan.....	11
3. Penilaian Keterampilan.....	13
F. <i>Pemodelan dalam Machine Learning</i>.....	14
1. Pembersihan Data.....	14
2. <i>Missing Data</i>	15
3. <i>Data Noisy</i>	16
4. <i>Outlier</i>	17
5. Metode untuk mengatasi <i>Outlier</i>	17
6. Variabel Nominal dan Variabel Ordinal.....	18

7.	<i>Ordinal Encoding</i>	19
8.	Transformasi Data	19
9.	Reduksi Data.....	20
BAB III METODOLOGI PENELITIAN		21
A. Tahapan Penelitian		21
1.	Identifikasi Masalah	22
2.	Perumusan Masalah dan Menetapkan Batasan Penelitian	22
3.	Pengumpulan Data	22
4.	<i>Pre-processing</i>	22
5.	<i>Clustering</i>	23
B. Data		23
C. Tools		32
D. Contoh / Sampel Data		32
BAB IV HASIL DAN PEMBAHASAN		36
A. Import Data		36
B. Pre - processing		37
1.	<i>Drop Data dan Identifikasi Missing Data</i>	37
2.	Mengatasi <i>Missing Data</i>	39
3.	Mengatasi Data Duplikat	40
4.	Mengatasi <i>Outlier</i>	41
5.	<i>Encoding Data</i>	44
6.	Normalisasi Data	47
7.	<i>Principal Component Anals (PCA)</i>	48
C. Clustering		53
D. Analisis Hasil Kluster		56
1.	Nilai rata – rata pengetahuan.....	57
2.	Nilai rata – rata Keterampilan	60
3.	Nilai Rata – rata KI1 dan KI2.....	63
4.	Analisis dengan Faktor Lain.....	68
5.	Persebaran Kelas per Kluster	77
BAB V KESIMPULAN DAN SARAN		81
A. Kesimpulan		81
B. Saran		82
DAFTAR PUSTAKA		83
LAMPIRAN		78
Lampiran 1. Beberapa Data Siswa Sekolah Dasar		87

Lampiran 2. <i>Source Code Pre – Processing Data</i>	88
Lampiran 3. <i>Source Code K-Means Clustering</i>	92
Lampiran 4. Eigen Vektor	93
Lampiran 5. Siswa Kuster 1	95
Lampiran 6. Siswa Kluster 2	97
Lampiran 7. Siswa Kluster 3	99
RIWAYAT HIDUP	101



DAFTAR TABEL

Tabel 1. Keterangan Atribut Data	24
Tabel 2. Contoh Data.....	32
Tabel 3. Hasil Identifikasi Missing Data	38
Tabel 4. Tabel Perhitungan Outlier	43
Tabel 5. Tabel Perhitungan Setelah Outlier Dihapus	44
Tabel 6. Sampel Data Sebelum Label Encoding	45
Tabel 7. Sampel Data Setelah Label Encoding	46
Tabel 8. Sampel Data sebelum dilakukan Ordinal Encoding.....	46
Tabel 9. Sampel Data setelah dilakukan Ordinal Encoding.....	47
Tabel 10. Akumulasi Rasio Varians	51
Tabel 11. Perhitungan Jarak pada SSE	54
Tabel 12. Kesimpulan Labeling Analisis Kluster.....	66
Tabel 13. Nama Kluster	67



DAFTAR GAMBAR

Gambar 1. Diagram alir algoritma K-Means Clustering.....	8
Gambar 2. Flowchart Metode Elbow.....	9
Gambar 3. Flowchart PCA.....	10
Gambar 4. Skema Penilaian Sikap	11
Gambar 5. Skema Penilaian Pengetahuan.....	12
Gambar 6. Skema Penilaian Keterampilan	14
Gambar 7. Flowchart Tahap Penelitian	21
Gambar 8. Informasi Data Bagian 1	36
Gambar 9. Informasi Data Bagian 2	37
Gambar 10. Grafik Identifikasi Missing Data.....	39
Gambar 11. Grafik Setelah Missing Data Kategorikal Diatasi	40
Gambar 12. Identifikasi Data Duplikat.....	41
Gambar 13. Sampel Data Setelah Outlier	42
Gambar 14. Sampel Data Setelah Outlier Bagian 2.....	42
Gambar 15. Sampel Data Setelah di-normalisasi	48
Gambar 16. Matriks Kovarian.....	49
Gambar 17. Nilai Eigen	49
Gambar 18. Eigen Vektor	50
Gambar 19. Varians PCA	50
Gambar 20. Grafik Kumulatif Persentase Varians PCA.....	51
Gambar 21. Hasil Reduksi Data dengan PCA.....	52
Gambar 22. Hasil Reduksi Data dengan PCA Bagian 2	53
Gambar 23. Hasil Elbow Method.....	54
Gambar 24. Hasil Clustering Menggunakan K-Means.....	55
Gambar 25. Grafik Persebaran Cluster	56
Gambar 26. Grafik Rata – rata Nilai Pengetahuan Siswa Kluster 1	57
Gambar 27. Grafik Rata – rata Nilai Pengetahuan Siswa Kluster 2	58
Gambar 28. Grafik Rata – rata Nilai Pengetahuan Siswa Kluster 3	58
Gambar 29. Rata – rata Pengetahuan Seluruh Kluster.....	59
Gambar 30. Grafik Rata – rata Nilai Keterampilan Siswa Kluster 1	60
Gambar 31. Grafik Rata – rata Nilai Keterampilan Siswa Kluster 2.....	61
Gambar 32. Grafik Rata – rata Nilai Keterampilan Siswa Kluster 2.....	62
Gambar 33. Rata – rata Keterampilan Seluruh Kluster	62
Gambar 34. Grafik Rata – rata Nilai KI1 dan KI2 Kluster 1	64
Gambar 35. Grafik Rata – rata Nilai KI1 dan KI2 Kluster 2	64
Gambar 36. Grafik Rata – rata Nilai KI1 dan KI2 Kluster 3	65
Gambar 37. Rata – rata KI1 & KI2 Seluruh Kluster	65
Gambar 38. Persebaran Jenis Kelamin Siswa Kluster Kreatif	69
Gambar 39. Persebaran Jenis Kelamin Siswa Kluster Progresif	69
Gambar 40. Persebaran Jenis Kelamin Siswa Kluster Intelektual.....	70
Gambar 41. Pekerjaan Ayah Siswa Kluster Kreatif	70
Gambar 42. Pekerjaan Ayah Siswa Kluster Progresif	71
Gambar 43. Pekerjaan Ayah Siswa Kluster Intelektual.....	71

Gambar 44. Pekerjaan Ibu Siswa Kluster Kreatif.....	72
Gambar 45. Pekerjaan Ibu Siswa Kluster Progresif	72
Gambar 46. Pekerjaan Ibu Siswa Kluster Intelektual	73
Gambar 47. Persebaran Jarak Rumah Kluster Kreatif	73
Gambar 48. Persebaran Jarak Rumah Kluster Progresif.....	74
Gambar 49. Persebaran Jarak Rumah Kluster Intelektual.....	74
Gambar 50. Persebaran Ekskul Kluster Kreatif	75
Gambar 51. Persebaran Ekskul Kluster Progresif	76
Gambar 52. Persebaran Ekskul Kluster Intelektual	76
Gambar 53. Total Siswa per Kelas	77
Gambar 54. Persebaran Kelas Siswa pada Kluster Kreatif	78
Gambar 55. Persebaran Kelas Siswa pada Kluster Progresif	78
Gambar 56. Persebaran Kelas Siswa pada Kluster Intelektual.....	79



DAFTAR LAMPIRAN

Lampiran 1. Beberapa Data Siswa Sekolah Dasar	87
Lampiran 2. <i>Source Code Pre – Processing Data</i>	88
Lampiran 3. <i>Source Code K-Means Clustering</i>	92
Lampiran 4. Eigen Vektor	93
Lampiran 5. Siswa Kuster 1.....	95
Lampiran 6. Siswa Kluster 2.....	97
Lampiran 7. Siswa Kluster 3.....	99





KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS NEGERI JAKARTA
UPT PERPUSTAKAAN

Jalan Rawamangun Muka Jakarta 13220
Telepon/Faksimili: 021-4894221
Laman: lib.unj.ac.id

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika Universitas Negeri Jakarta, yang bertanda tangan di bawah ini, saya:

Nama : Rachel Haryawan
NIM : 1313618026
Fakultas/Prodi : Fakultas Matematika dan Ilmu Pengetahuan Alam / Ilmu Komputer.
Alamat email : rachelharyawan@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada UPT Perpustakaan Universitas Negeri Jakarta, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah:

Skripsi Tesis Disertasi Lain-lain (.....)

yang berjudul :

Klusterisasi Siswa Sekolah Dasar dengan
Algoritma K-Means Clustering

Dengan Hak Bebas Royalti Non-Eksklusif ini UPT Perpustakaan Universitas Negeri Jakarta berhak menyimpan, mengalihmediakan, mengelolanya dalam bentuk pangkalan data (*database*), mendistribusikannya, dan menampilkan/mempublikasikannya di internet atau media lain secara *fulltext* untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan Universitas Negeri Jakarta, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 30 Agustus 2023

Penulis

(Rachel Haryawan)
nama dan tanda tangan