

**PENERAPAN ALGORITMA K-MEANS UNTUK
PENGELOMPOKAN NILAI UJIAN NASIONAL DI SDN
GEDANGAN 02 SUKOHARJO**

SKRIPSI



disusun oleh

Andhika Putratama

18.11.2059

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2022**

**PENERAPAN ALGORITMA K-MEANS UNTUK
PENGELOMPOKAN NILAI UJIAN NASIONAL DI SDN
GEDANGAN 02 SUKOHARJO**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh

Andhika Putratama

18.11.2059

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2022**

PERSETUJUAN

SKRIPSI

**PENERAPAN ALGORITMA K-MEANS UNTUK
PENGELOMPOKAN NILAI UJIAN NASIONAL
DI SDN GEDANGAN 02 SUKOHARJO**

yang dipersiapkan dan disusun oleh

Andhika Putratama

18.11.2059

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 20 Juli 2022

Dosen Pembimbing,

Bayu Setiaji, M.Kom

NIK. 190302116

PENGESAHAN
SKRIPSI
PENERAPAN ALGORITMA K-MEANS UNTUK
PENGELOMPOKAN NILAI UJIAN NASIONAL
DI SDN GEDANGAN 02 SUKOHARJO

yang dipersiapkan dan disusun oleh

Andhika Putratama

18.11.2059

telah dipertahankan di depan Dewan Penguji
pada tanggal 20 Juli 2022

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Krisnawati, S.Si., M.T.
NIK. 190302038

Mardhiya Hayatv, S.T., M.Kom
NIK. 190302108

Bayu Setiaji, M.Kom
NIK. 190302216

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 20 Juli 2022

DEKAN FAKULTAS ILMU KOMPUTER

Hanif Al Fatta, S.Kom., M.Kom.
NIK. 190302096

PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Yogyakarta, 20 Juli 2022



Andhika Putratama

NIM. 18.11.2059

MOTTO

*ORANG YANG HEBAT ADALAH ORANG YANG MEMILIKI KEMAMPUAN
MENYEMBUNYIKAN KESUSAHAN, SHINGGA ORANG LAIN MENGIRA BAHWA IA
SELALU SENANG*

[IMAM SYAFI'I]



PERSEMBAHAN

Alhamdulillah Puji Syukur ini penulis panjatkan, akhirnya skripsi ini dapat terselesaikan dengan baik. Selaku penulis mempersembahkan skripsi ini kepada:

1. Allah SWT atas ridho-Nya, skripsi ini dapat terselesaikan dengan baik, Sujud syukur saya panjatkan kepada-Mu.
2. Kedua orang tua saya yang telah banyak berjasa dalam hidup saya, selalu memberikan semangat untuk tetap menuntut ilmu dan selalu memberikan motivasi untuk menyelesaikan skripsi ini.
3. Keluarga besarku, kakak-kakakku, keponakanku, terima kasih telah menjadi penyemangat untuk menjadikan ku sosok yang kuat dan tabah dalam menjalani kehidupan ini.
4. Teruntuk Bapak Bayu Setiaji, M.Kom selaku dosen pembimbing, saya ucapkan terima kasih karena telah bersedia membimbing, selalu memberikan kritik dan saran yang membangun serta mengantarkan saya untuk mengantungi gelar sarjana.
5. Teman-teman 18 IF 04, terima kasih banyak untuk bantuan dan kerja samanya selama ini, teman seperjuangan saya selama ini.
6. Serta semua pihak yang tidak bias saya sebutkan satu persatu yang telah mendoakan, mendukung dan memotivasi saya selama ini.

KATA PENGANTAR

Dengan memanjatkan puja dan puji syukur kehadirat Allah SWT yang telah melimpahkan rahmat, taufik dan hidayah-Nya, sehingga penulis dapat menyelesaikan laporan skripsi ini dengan judul **“Penerapan Algoritma K-Means Untuk Pengelompokan Nilai Ujian Nasional Di SDN Gedangan 02 Sukoharjo”**, sebagai salah satu syarat untuk menyelesaikan Program Sarjana (S1) Jurusan Informatika Universitas Amikom Yogyakarta.

Penulis menyadari bahwa skripsi ini tidak mungkin terselesaikan tanpa adanya dukungan, bantuan, bimbingan, dan nasehat dari berbagai pihak selama penyusunan skripsi ini. Pada kesempatan ini, penulis menyampaikan terima kasih setulus-tulusnya kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M selaku Rektor Universitas Amikom Yogyakarta.
2. Bapak Hanif Al Fatta selaku Dekan Fakultas Ilmu Komputer
3. Ibu Windha Mega Pradnya Dhuhita, M.Kom selaku Ketua Program Studi Informatika di Universitas Amikom Yogyakarta.
4. Bapak Bayu Setiaji M.Kom, sebagai dosen pembimbing yang telah membantu saya dalam memberikan petunjuk, saran, pengarahan, serta bimbingan yang besar dalam skripsi ini.
5. Kedua orang tua penulis, yang selalu memberikan kasih sayang, doa, nasehat, serta kesabarannya yang luar biasa dalam setiap langkah hidup

penulis, yang banyak memberikan dukungan moril serta materil dengan tulus ikhlas.

6. Teman-teman yang telah ikut andil dan banyak membantu dalam penyelesaian skripsi ini.
7. Seluruh staff dan karyawan Universitas Amikom Yogyakarta yang banyak membantu kelancaran segala aktivitas penulisan skripsi ini.

Penulis menyadari bahwa dalam penulisan skripsi ini masih banyak kekurangan kritik dan saran yang membangun selalu penulis harapkan demi kemajuan dan menuju arah yang lebih baik di masa yang akan datang sehingga dapat bermanfaat bagi penulis serta pihak-pihak yang membutuhkan. Semoga laporan skripsi ini bermanfaat bagi semua pihak.

Yogyakarta, 20 Juli 2022

Andhika Putratama

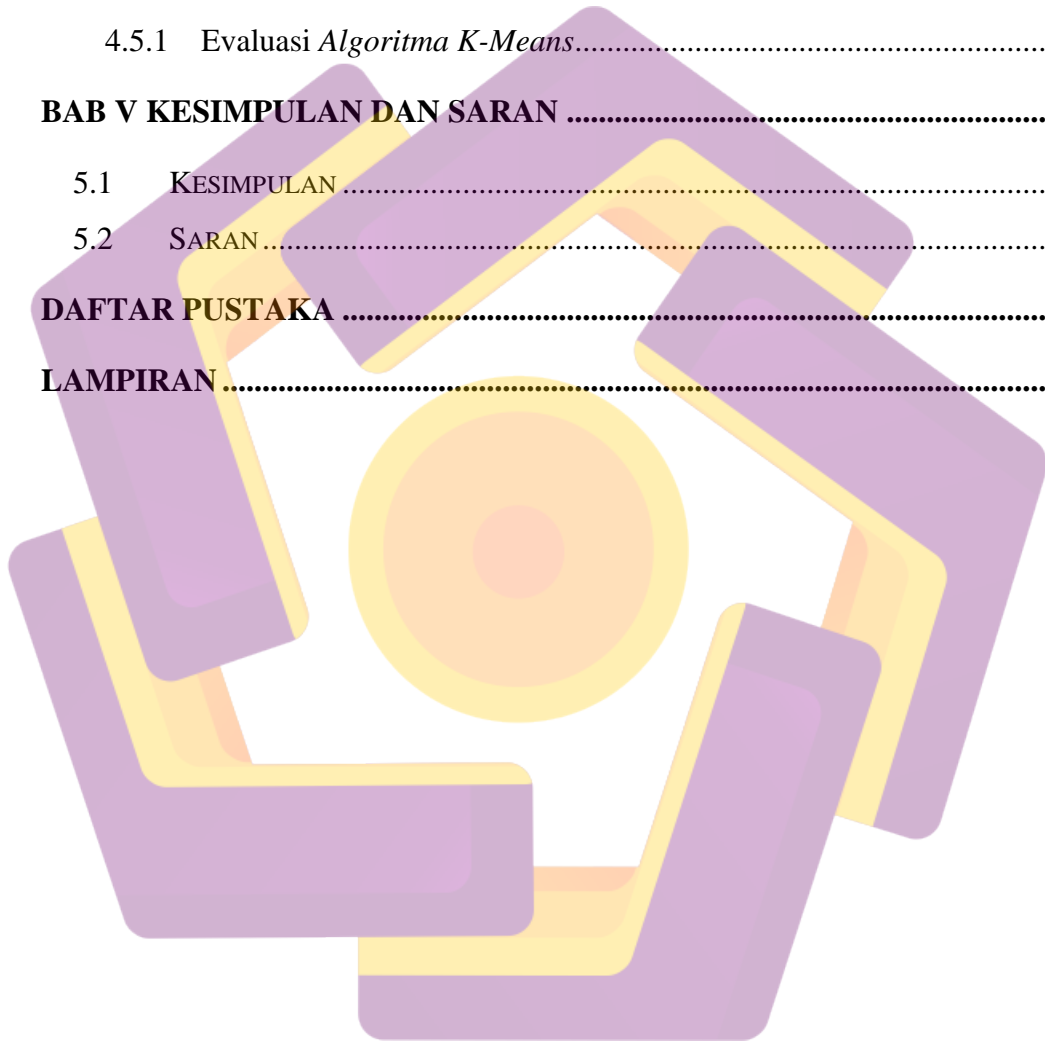
18.11.2059

DAFTAR ISI

HALAMAN JUDUL	I
PPERSETUJUAN	II
PERSETUJUAN	II
PENGESAHAN	III
PENGESAHAN	III
PERNYATAAN	IV
MOTTO	V
PERSEMBAHAN	VI
KATA PENGANTAR	VII
DAFTAR ISI	IX
DAFTAR TABEL	XII
DAFTAR GAMBAR	XIII
INTISARI	XIV
ABSTRACT	XV
BAB I PENDAHULUAN	1
1.1 LATAR BELAKANG	1
1.2 RUMUSAN MASALAH	3
1.3 BATASAN MASALAH	3
1.4 MAKSUD DAN TUJUAN PENELITIAN	4
1.5 MANFAAT PENELITIAN	4
1.6 METODE PENELITIAN	5
1.6.1 Metode Pengumpulan Data	5
1.6.2 Metode Analisis	5
1.7 SISTEMATIKA PENULISAN	6

BAB II LANDASAN TEORI.....	8
2.1 TINJAUAN PUSTAKA	8
2.2 <i>DATA MINING</i>	14
2.2.1 Fungsi Data Mining	16
2.3 <i>CLUSTERING</i>	18
2.4 <i>ALGORITMA K-MEANS</i>	18
2.5 <i>ROBUST SCALER</i>	21
2.6 <i>MINMAX SCALER</i>	22
2.7 PENGUJIAN TINGKAT <i>CLUSTERISASI</i>	22
2.7.1 <i>Metode Elbow</i>	22
2.8 UJIAN NASIONAL	23
BAB III METODOLOGI PENELITIAN.....	25
3.1 ALAT DAN BAHAN	25
3.1.1 Perangkat Keras	25
3.1.2 Perangkat Lunak	26
3.2 TAHAPAN PENELITIAN	26
3.2.1 Identifikasi Masalah	27
3.2.2 Studi Literatur	28
3.2.3 <i>Selection Data</i>	28
3.2.4 <i>Data Cleaning</i>	29
3.2.5 <i>Data Reduction</i>	34
3.2.6 <i>Data Transformation</i>	34
3.2.7 <i>Data Mining</i>	42
3.2.8 <i>Evaluation</i>	57
BAB IV HASIL DAN PEMBAHASAN.....	58
4.1 IMPLEMENTASI <i>CLUSTERING K-MEANS</i>	58
a. Mendapatkan data	58
b. <i>Data Cleaning</i>	59
c. <i>Feature Selection</i>	60
d. <i>Data Reduction</i>	61
e. <i>Data Transformation</i>	61

f. <i>Clustering K-means</i>	67
4.2 PENGUJIAN	70
4.3 PROSEDUR PENGUJIAN	71
4.3.1 <i>Prosedur Pengujian Algoritma K-Means</i>	71
4.4 DATA HASIL PENGUJIAN	71
4.4.1 <i>Hasil Pengujian Algoritma K-Means</i>	71
4.5 ANALISA DATA / EVALUASI	74
4.5.1 <i>Evaluasi Algoritma K-Means</i>	74
BAB V KESIMPULAN DAN SARAN	75
5.1 KESIMPULAN	75
5.2 SARAN	76
DAFTAR PUSTAKA	77
LAMPIRAN	81



DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	10
Tabel 3.1 Data Awal	29
Tabel 3.2 Perhitungan Rata-rata dan Standar Deviasi.....	33
Tabel 3.3 Perhitungan Threshold	33
Tabel 3.4 Hasil Transformasi Bahasa Indonesia	35
Tabel 3.5 Hasil Transformasi Matematika	38
Tabel 3.6 Hasil Transformasi IPA	41
Tabel 3.7 Titik Pusat Awal	43
Tabel 3.8 Hasil Perhitungan Euclidean Distance	44
Tabel 3.9 Titik Pusat Iterasi 2	45
Tabel 3.10 Hasil Iterasi 2	46
Tabel 3.11 Titik Pusat Iterasi 3	47
Tabel 3.12 Hasil Iterasi 3	48
Tabel 3.13 Titik Pusat Iterasi 4	49
Tabel 3.14 Hasil Iterasi 4	50
Tabel 3.15 Titik Pusat Iterasi 5	51
Tabel 3.16 Hasil Iterasi 5	52
Tabel 3.17 Titik Pusat Iterasi 6	53
Tabel 3.18 Hasil Iterasi 6	54
Tabel 3.19 Titik Pusat Iterasi 7	55
Tabel 3.20 Hasil Iterasi 7	56
Tabel 4.1 Hasil Transformasi Setiap Atribut.....	66
Tabel 4.2 Titik Pusat Awal	68
Tabel 4.3 Hasil Clustering K-Means.....	68

DAFTAR GAMBAR

Gambar 2.1 Proses Knowledge Discovery in Database (KDD)[12]	14
Gambar 2.2 Fungsi Data Mining[13].....	16
Gambar 2.3 Bagan Alur Algoritma Clustering K-Means[19].....	19
Gambar 3.1 Tahapan Penelitian	27
Gambar 3.2 Histogram B.Indonesia Sebelum Transformasi	34
Gambar 3.3 Histogram Bahasa Indonesia Setelah Transformasi.....	35
Gambar 3.4 Histogram Matematika Sebelum Transformasi.....	37
Gambar 3.5 Histogram Matematika Sesudah Transformasi	38
Gambar 3.6 Histogram IPA Sebelum Transformasi.....	40
Gambar 3.7 Hsistogram IPA Setelah Transformasi.....	40
Gambar 4.1 Source Code pengambilan data	58
Gambar 4.2 Source Code Visualisasi Atribut.....	59
Gambar 4.3 Source Code Data Cleaning	59
Gambar 4.4 Source Code Feature Selection	60
Gambar 4.5 Grafik Heatmap Person Correlation Matrix.....	60
Gambar 4.6 Source Code Data Reduction	61
Gambar 4.7 Source Code Data Transformation	61
Gambar 4.8 Source Code Transformasi Bahasa Indonesia.....	62
Gambar 4.9 Boxplot Bahasa Indonesia	63
Gambar 4.10 Source Code Transformasi Matematika.....	63
Gambar 4.11 Boxplot Matematika	64
Gambar 4.12 Source Code Transformasi IPA.....	64
Gambar 4.13 Boxplot IPA.....	65
Gambar 4.14 Source Code Clustering K-Means	67
Gambar 4.15 Visualisasi Subplot Clustering K-Means	70
Gambar 4.16 Source Code Elbow Method.....	72
Gambar 4.17 Hasil Uji Elbow Method	73

INTISARI

Di dalam Penilaian Pendidikan, Ujian Nasional (UN) sangatlah penting, karena Ujian Nasional (UN) merupakan evaluasi standar pendidikan yang diadakan setiap tahunnya. Ujian Nasional (UN) merupakan suatu pencapaian kompetensi lulusan secara nasional pada mata pelajaran tertentu pada Sekolah Dasar (SD), Sekolah Menengah Pertama (SMP), Sekolah Menengah Atas (SMA) / Sekolah Menengah Kejuruan (SMK). Berdasarkan data hasil Ujian Nasional (UN) yang disajikan di website resmi Kementerian Pendidikan dan Kebudayaan (KEMDIKBUD), data yang disajikan tidak berdasarkan peringkat atau kategori tertentu sehingga informasi yang didapatkan kurang jelas sehingga masyarakat atau orang tua tidak memiliki acuan saat mendaftarkan putra-putri mereka ke jenjang berikut.

Hal itu juga terjadi pada SDN Gedangan 02 Sukoharjo, dimana orang tua tidak mendapatkan informasi berkaitan dengan peringkat atau kategori nilai Ujian Nasional terhadap putra putrinya. Dengan adanya permasalahan tersebut, maka dilakukan sebuah penelitian dengan menggunakan data nilai rata-rata Ujian Nasional tingkat SD di SDN Gedangan 02 Sukoharjo menggunakan teknik data mining yaitu *algoritma k-means clustering* dengan mengelompokkan data ke dalam beberapa cluster berdasarkan kemiripan datanya. Berdasarkan hasil penelitian, teridentifikasi 3 *cluster* yaitu *cluster 0* dengan tingkat pengelompokan nilai Ujian Nasional kategori tinggi terdapat 5 siswa, *cluster 1* dengan tingkat pengelompokan nilai Ujian Nasional kategori sedang terdapat 16 siswa, *cluster 2* dengan tingkat pengelompokan nilai Ujian Nasional kategori rendah terdapat 4 siswa.

Berdasarkan hasil penelitian tersebut, diharapkan dapat memberikan informasi serta mendukung SDN Gedangan 02 Sukoharjo untuk melakukan pengelompokan nilai Ujian Nasional berdasarkan tiap *cluster*-nya

Kata Kunci: *Data Mining, Algoritma K-Means, Pengelompokan, Ujian Nasional*

ABSTRACT

In Educational Assessment, the National Examination (UN) is very important, because the National Examination (UN) is an evaluation of educational standards that is held every year. The National Examination (UN) is an achievement of national graduate competence in certain subjects at Elementary School (SD), Junior High School (SMP), High School (SMA) / Vocational High School (SMK). Based on the results of the National Examination (UN) presented on the official website of the Ministry of Education and Culture (KEMDIKBUD), the data presented is not based on certain rankings or categories so that the information obtained is not clear so that the public or parents do not have a reference when registering their children. to the next level.

This also happened at SDN Gedangan 02 Sukoharjo, where parents did not receive information related to the ranking or category of National Examination scores for their sons and daughters. Given these problems, a study was conducted using the data on the average value of the SD National Examination at SDN Gedangan 02 Sukoharjo using data mining techniques, namely the k-means clustering algorithm by grouping the data into several clusters based on the similarity of the data. Based on the results of the study, 3 clusters were identified, namely cluster 0 with a high level of grouping of National Examination scores, there were 5 students, cluster 1 with a grouping level of National Examination scores in the medium category, cluster 2 with a grouping level of National Examination scores in the low category, there were 4 students.

Based on the results of this study, it is hoped that it can provide information and support SDN Gedangan 02 Sukoharjo to group the National Examination scores based on each cluster.

Keyword: *Data Mining, K-Means Algorithm, Clustering, National Examination*