



**REDUKSI DIMENSI DENGAN *PRINCIPAL COMPONENT ANALYSIS*
PADA PENGELOMPOKAN DATA *TIME SERIES* SENSOR GAS
MENGUNAKAN *CLUSTERING LARGE APPLICATION (CLARA)***

SKRIPSI

Untuk Memenuhi Persyaratan
Dalam Menyelesaikan Sarjana Strata – 1 Ilmu Komputer

Oleh
FALAH ISNAIN
NIM 1911016210008

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
JUNI 2024**



**REDUKSI DIMENSI DENGAN *PRINCIPAL COMPONENT ANALYSIS* PADA
PENGELOMPOKAN DATA *TIME SERIES* SENSOR GAS MENGGUNAKAN
*CLUSTERING LARGE APPLICATION (CLARA)***

SKRIPSI

Untuk Memenuhi Persyaratan

Dalam Menyelesaikan Sarjana Strata – 1 Ilmu Komputer

Oleh

FALAH ISNAIN

NIM 1911016210008

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
JUNI 2024**

SKRIPSI

**REDUKSI DIMENSI DENGAN *PRINCIPAL COMPONENT ANALYSIS* PADA
PENGELOMPOKAN DATA *TIME SERIES* SENSOR GAS MENGGUNAKAN
*CLUSTERING LARGE APPLICATION (CLARA)***

Oleh

FALAH ISNAIN

1911016210008

Telah dipertahankan di depan Dosen Penguji pada tanggal 21 Juni 2024.

Susunan Dosen Penguji :

Pembimbing I



Dwi Kartini, S.Kom., M.Kom.

NIP. 198704212012122003

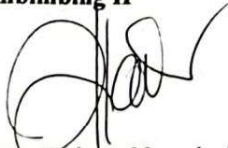
Dosen Penguji I



Irwan Budiman, S.T., M.Kom.

NIP. 197703252008121001

Pembimbing II



Dodon Turianto Nugrahadi, S.Kom, M.Eng.

NIP. 198001122009121002

Dosen Penguji II



Triando H. Saragih, S.Kom., M.Kom.

NIP. 199308242019031012



Banjarbaru, 1 Juli 2024

Program Studi Ilmu Komputer,

Irwan Budiman, S.T., M.Kom.

NIP. 197703252008121001

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam Daftar Pustaka.

Banjarbaru, 21 Juni 2024



FALAH ISNAIN

NIM. 1911016210008

ABSTRAK

REDUKSI DIMENSI DENGAN *PRINCIPAL COMPONENT ANALYSIS* PADA PENGELOMPOKAN DATA *TIME SERIES* SENSOR GAS MENGGUNAKAN *CLUSTERING LARGE APPLICATION (CLARA)*

(Oleh: Falah Isnain ; Pembimbing: Dwi Kartini, S.Kom., M.Kom.; Dodon Turianto Nugrahadi, S.Kom, M.Eng.; 2024; 56 halaman)

Data *time series* merupakan data dinamis yang berskala besar dengan dimensi tinggi yang kadang memiliki variabel yang *noise*. *Clustering* data *time series* dapat memberikan pola yang menarik dan potensi informasi di antara kumpulan data yang *homogen*. Namun *Multikolinearitas* sering menjadi masalah dalam melakukan *Clustering* dengan adanya hubungan yang mendekati linear diantara variabel-variabel. Korelasi yang tinggi antara dua variabel memiliki arti variabel memiliki tren yang sama dan cenderung membawa informasi yang sama. Pemeriksaan hubungan antar variabel sangat penting karena variabel yang sangat berkorelasi dapat mempengaruhi kinerja hasil *clustering*. Tujuan dari penelitian ini adalah untuk mengetahui apakah dengan menggunakan reduksi dimensi dan *clustering* berbasis partisi dapat meningkatkan hasil kinerja *clustering* yang telah ditentukan pada data *time series* emisi gas. Metode reduksi dimensi yang digunakan adalah *Principal component analysis*, dan metode *clustering* yang digunakan adalah CLARA. PCA digunakan untuk mengurangi korelasi antar variabel pada dataset dengan jumlah komponen yang telah ditentukan. Tujuan dari penelitian ini adalah melakukan perbandingan kinerja antara *clustering* berbasis partisi tanpa reduksi dimensi dan menggunakan reduksi dimensi. Hasil kinerja *clustering* terbaik diperoleh nilai *silhouette score coefficient* pada metode CLARA menggunakan reduksi dimensi PCA menjadi 2 komponen sebesar 0.4904 dengan jumlah *cluster* 2, serta nilai *Davies Bouldin Index (DBI)* terbaik pada metode CLARA menggunakan reduksi dimensi PCA menjadi 2 komponen sebesar 0.4904 dengan jumlah *cluster* 2.

Kata kunci: Reduksi Dimensi, *Principal Component Analysis*, *Clustering*, CLARA.

ABSTRACT

DIMENSION REDUCTION WITH PRINCIPAL COMPONENT ANALYSIS ON CLUSTERING TURBINE ENGINE CO AND NOX GAS EMISSION DATA USING CLUSTERING LARGE APPLICATION (CLARA)

(By: Falah Isnain; Advisors: Dwi Kartini, S.Kom., M.Kom.; Dodon Turianto Nugrahadi, S.Kom, M.Eng.; 2024; 56 pages)

Time series data is dynamic data that is large-scale, high-dimensional data that sometimes has variable noise. Clustering time series data can provide interesting patterns and potential information among homogeneous data sets. However, multicollinearity is often a problem in Clustering with the presence of a near-linear relationship between the variables. A high correlation between two variables means that the variables have similar trends and tend to carry the same information. Checking the relationship between variables is very important because highly correlated variables can affect the performance of Clustering results. The purpose of this research is to find out whether using dimensionality reduction and partition-based Clustering can improve the performance of Clustering results. dimensionality reduction and partition-based Clustering can improve the results of the predetermined Clustering performance on the data. performance on gas emission time series data. The dimension reduction method used is Principal component analysis, and the Clustering method used is CLARA. the Clustering method used is CLARA. PCA is used to reduce correlation between variables in the dataset with a predetermined number of components. determined. The purpose of this research is to compare the performance performance between partition-based Clustering without dimensionality reduction and using dimensionality reduction. dimension reduction. The best Clustering performance results are obtained by the silhouette value score coefficient in the CLARA method using dimensional reduction of PCA to 2 component of 0.4904 with the number of clusters 2, as well as the best Davies Bouldin Index (DBI) value in the CLARA method. Index (DBI) value is best in the CLARA method using PCA dimensional reduction to 2 components of 0.4904 with the number of clusters 2.

Keywords: Dimension Reduction, Principal Component Analysis, *Clustering*, CLARA.

PRAKATA

Puji dan syukur kepada ALLAH SWT atas berkat rahmat-Nya penulis dapat menyelesaikan skripsi yang berjudul **REDUKSI DIMENSI DENGAN *PRINCIPAL COMPONENT ANALYSIS* PADA PENGELOMPOKAN DATA *TIME SERIES* SENSOR GAS MENGGUNAKAN *CLUSTERING LARGE APPLICATION* (CLARA)** untuk memenuhi syarat dalam menyelesaikan pendidikan program Strata-1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat.

Tidak lupa penulis menyampaikan ucapan terimakasih kepada pihak-pihak terkait yang sangat mendukung dan membantu penulis dalam penyusunan skripsi ini, adapun yang dimaksud adalah sebagai berikut:

1. Allah SWT, karena atas limpahan karunia-Nya lah penulis dapat menyelesaikan skripsi ini.
2. Ayah, Ibu, dan seluruh keluarga yang selalu memberikan doa, semangat, dukungan, hingga kepercayaan yang membuat penulis selalu bekerja keras menyelesaikan skripsi.
3. Ibu Dwi Kartini, S. Kom., M. Kom. dan Dodon Turianto Nugrahadi, S.Kom, M.Eng. selaku dosen pembimbing utama dan pendamping pada pembuatan skripsi ini.
4. Irwan Budiman, S.T., M.Kom. dan Triando H. Saragih, S.Kom., M.Kom selaku dosen ketua penguji dan anggota penguji dalam skripsi ini.
5. Bapak Irwan Budiman, S.T., M.Kom. selaku ketua program studi Ilmu Komputer beserta seluruh dosen dan karyawan/staff pegawai Fakultas MIPA Universitas Lambung Mangkurat atas bantuan yang diberikan selama penulis mengikuti studi.
6. Raditya Raka Alif Azizi dan Khairi Yahya yang telah memfasilitasi dan membantu dengan segala kritik, saran dan masukan. dalam menjalankan penelitian.
7. Muhammad Thoriq hidayat, Muhammad Meftah Mafazy, Muhammad Hanan Ababil, M. Fajar Aditya, Bagas Rony Hidayat, Naufal Abiyyu Mayusuf, Dwa

Meizadewa, Muhammad Nouval Rahman, Muhammad Farros Zulfan, dan Muhammad Khairie. Sebagai kawan dalam berkeluh kesah, bertukar fikiran, menjalankan perkuliahan, serta tempat mengembangkan diri.

8. Intelcore 2019 yang telah membantu dalam menjalankan perkuliahan bersama dari awal sampai akhir.
9. Seluruh mahasiswa Ilmu Komputer ULM yang telah membantu dalam keseharian di lingkungan kampus ULM Banjarbaru.
10. Semua pihak yang tidak dapat disebutkan satu persatu yang telah turut membantu dalam menyelesaikan skripsi ini.

Akhir kata penulis menyadari bahwa skripsi ini jauh dari sempurna, kritik dan saran yang membangun sangat diperlukan. Semoga tulisan ini dapat bermanfaat bagi ilmu pengetahuan dan pembaca khususnya serta mendapat keridhaan Allah SWT.

Banjarbaru, 21 Juni 2024



Falah Isnain

DAFTAR ISI

PERNYATAAN	iv
ABSTRAK	v
ABSTRACT	vi
PRAKATA	vii
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN	xiv
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Tujuan.....	3
1.4. Manfaat Penelitian.....	3
1.5. Batasan Masalah.....	3
BAB II TINJAUAN PUSTAKA	5
2.1 Literatur Terdahulu	5
2.2 <i>Machine Learning</i>	7
2.3 Normalisasi <i>Z-Score</i>	7
2.4 <i>Principal Component Analysis(PCA)</i>	8
2.5 <i>Metode Clustering</i>	9
2.6 <i>Clustering Large Application (CLARA)</i>	10
2.7 <i>Silhouette Coefficient</i>	11
2.8 <i>Davies Bouldin Index (DBI)</i>	12
2.9 Keaslian Penelitian	14
BAB III METODE PENELITIAN	18
3.1 Alat Penelitian	18
3.2 Bahan Penelitian	18

3.3	Prosedur Penelitian.....	19
a)	Silhouette Coefficient.....	20
b)	Davies Bouldin Index (DBI)	20
BAB IV HASIL DAN PEMBAHASAN		21
4.1	Hasil	21
4.1.1	Pengumpulan Data	21
4.1.2	Preprocessing.....	22
4.1.3	<i>Principal Component Analysis (PCA)</i>	24
4.1.4	<i>Clustering Large Application (CLARA)</i>	31
4.1.5	PCA - CLARA	35
4.1.6	Evaluasi Menggunakan Metode <i>Davies Bouldin Index (DBI)</i>	40
4.1.7	Evaluasi Menggunakan Metode Silhouette Score	45
4.2	Pembahasan.....	48
BAB V PENUTUP.....		54
5.1	Kesimpulan.....	54
5.2	Saran.....	54
DAFTAR PUSTAKA		55
LAMPIRAN.....		58

DAFTAR TABEL

Tabel 1 Keaslian Penelitian.....	14
Tabel 2 Perancangan Penelitian	17
Tabel 3 Nama fitur	21
Tabel 4 Contoh <i>Turbin gase CO and NOx Emission Dataset</i>	22
Tabel 5 Perhitungan <i>Z-Score</i>	23
Tabel 6. Hasil Perhitungan <i>Z-Score</i>	24
Tabel 7. Hasil Perhitungan <i>Covariance</i>	24
Tabel 8 Hasil Perhitungan <i>Eigenvalue</i>	25
Tabel 9. Hasil perhitungan <i>eigenvector</i>	25
Tabel 10. Perhitungan nilai <i>principal component</i>	26
Tabel 11. Hasil reduksi dimensi PCA.....	26
Tabel 12. Contoh data yang sudah di normalisasi.....	27
Tabel 13 hasil perhitungan matrix <i>Covariance</i>	28
Tabel 14 Hasil contoh reduksi dimensi PCA	30
Tabel 15 <i>Data gas emission CO and NOx</i> yang sudah dinormalisasi	31
Tabel 16 Pusat <i>medoid</i> awal CLARA	31
Tabel 17 Hasil CLARA iterasi 1	32
Tabel 18 Pusat <i>medoid</i> iterasi 2	32
Tabel 19 Hasil CLARA iterasi 2.....	33
Tabel 20 Pengelompokan CLARA dengan $K = 2$	34
Tabel 21 Hasil pengelompokan CLARA dengan $K = 3$	34
Tabel 22 Hasil pengelompokan CLARA dengan $K = 4$	34
Tabel 23 Hasil pengelompokan CLARA dengan $K = 5$	34
Tabel 24 Hasil pengelompokan CLARA dengan $K = 6$	34
Tabel 25 Hasil pengelompokan CLARA dengan $K = 7$	35
Tabel 26 Hasil pengelompokan CLARA dengan $K = 8$	35
Tabel 27 Hasil pengelompokan CLARA dengan $K = 9$	35
Tabel 28 Hasil pengelompokan CLARA dengan $K = 10$	35
Tabel 29 Sampel data yang sudah direduksi.....	35

Tabel 30 Pusat <i>medoid</i> awal PCA-CLARA	36
Tabel 31 Hasil PCA-CLARA iterasi 1	36
Tabel 32 Pusat <i>Medoid</i> iterasi 2 PCA-CLARA	37
Tabel 33 Hasil PCA-CLARA iterasi 2	37
Tabel 34 Pengelompokan PCA-CLARA dengan $K = 2$	38
Tabel 35 Hasil pengelompokan PCA-CLARA dengan $K = 3$	38
Tabel 36 Hasil pengelompokan PCA-CLARA dengan $K = 4$	39
Tabel 37 Hasil pengelompokan PCA-CLARA dengan $K = 5$	39
Tabel 38 Hasil pengelompokan PCA-CLARA dengan $K = 6$	39
Tabel 39 Hasil pengelompokan PCA-CLARA dengan $K = 7$	39
Tabel 40 Hasil pengelompokan PCA-CLARA dengan $K = 8$	39
Tabel 41 Hasil pengelompokan PCA-CLARA dengan $K = 9$	40
Tabel 42 Hasil pengelompokan PCA-CLARA dengan $K = 10$	40
Tabel 43 Perhitungan SSW <i>Cluster 1</i> CLARA.....	40
Tabel 44 Perhitungan SSW <i>Cluster 2</i> CLARA.....	41
Tabel 45 Tabel perhitungan SSB CLARA.....	41
Tabel 46 Perhitungan Rasio CLARA.....	42
Tabel 47 Perhitungan SSW <i>Cluster 1</i> PCA-CLARA.....	42
Tabel 48 Perhitungan SSW <i>Cluster 2</i> PCA-CLARA.....	43
Tabel 49 Perhitungan SSB PCA-CLARA.....	43
Tabel 50 Perhitungan Rasio PCA-CLARA	44
Tabel 51. Hasil Evaluasi DBI.....	44
Tabel 52 Tabel hasil perhitungan <i>silhouette coefficient</i> CLARA	46
Tabel 53 Tabel hasil perhitungan <i>silhouette coefficient</i> PCA-CLARA.....	47
Tabel 54. Hasil Evaluasi <i>Silhouutte Coefficient</i>	47
Tabel 55 Hasil terbaik <i>Silhouette Coefficient</i>	52
Tabel 56 Hasil terbaik <i>Davies Bouldin Index</i> (DBI).....	52

DAFTAR GAMBAR

Gambar 1 Prosedur Penelitian.....	19
Gambar 2 Korelasi pada data	23
Gambar 3 Hasil <i>Silhouette Coefficient</i> CLARA	49
Gambar 4 Hasil DBI CLARA	49
Gambar 5 Hasil <i>Silhouette Coefficient</i> CLARA PC2	50
Gambar 6 Hasil DBI CLARA PC2	50
Gambar 9 Perbandingan <i>Silhouette Coefficient</i>	51
Gambar 10 Perbandingan <i>Davies Bouldin Index</i> (DBI).....	52

DAFTAR LAMPIRAN

Lampiran 1 Code Untuk Mengimpor Library Yang Digunakan.....	59
Lampiran 2 Code Untuk Import dataset.....	59
Lampiran 3 Code Untuk CLARA <i>Clustering</i> dengan jarak <i>manhattan</i>	59
Lampiran 4 Code Untuk CLARA <i>Clustering</i> dengan jarak <i>euclidean</i>	60
Lampiran 5 Code Untuk Evaluasi Silhouette Coefficient CLARA <i>Manhattan</i>	61
Lampiran 6 Code Untuk Evaluasi Silhouette Coefficient CLARA <i>Euclidean</i>	61
Lampiran 7 Code Untuk Evaluasi Davies Bouldin Indeks CLARA <i>Manhattan</i>	61
Lampiran 8 Code Untuk Evaluasi Davies Bouldin Indeks CLARA <i>Euclidean</i>	62
Lampiran 9 Code Untuk Reduksi Dimensi PCA	63
Lampiran 10 Code Untuk CLARA <i>Manhattan</i> dan PCA	63
Lampiran 11 Code Untuk CLARA <i>Euclidean</i> dan PCA	64
Lampiran 12 Code Untuk <i>Silhouette score coefficient</i> CLARA <i>Manhattan</i> dan PCA	65
Lampiran 13 Code Untuk <i>Silhouette score coefficient</i> CLARA <i>Euclidean</i> dan PCA	65
Lampiran 14 Code Untuk <i>Davies Bouldin Index</i> CLARA <i>Manhattan</i> dan PCA	65
Lampiran 15 Code Untuk <i>Davies Bouldin Index</i> CLARA <i>Euclidean</i> dan PCA.....	66
Lampiran 16 Riwayat Hidup Penulis	68