



**PERBANDINGAN PRINCIPAL COMPONENT ANALYSIS (PCA) DAN
SPECTRAL CLUSTERING DALAM PENGELOMPOKAN DATA
MICROARRAY MENGGUNAKAN K-MEANS**

SKRIPSI

Untuk Memenuhi Persyaratan

Dalam Menyelesaikan Sarjana Strata – 1 Ilmu Komputer

Oleh

RADITYA RAKA ALIF AZIZI

NIM 1911016110001

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



**PERBANDINGAN PRINCIPAL COMPONENT ANALYSIS (PCA) DAN
SPECTRAL CLUSTERING DALAM PENGELOMPOKKAN DATA
MICROARRAY MENGGUNAKAN K-MEANS**

SKRIPSI

Untuk Memenuhi Persyaratan

Dalam Menyelesaikan Sarjana Strata – 1 Ilmu Komputer

Oleh

RADITYA RAKA ALIF AZIZI

NIM 1911016110001

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

SKRIPSI
PERBANDINGAN PRINCIPAL COMPONENT ANALYSIS (PCA)
DENGAN SPECTRAL CLUSTERING DALAM PENGELOMPOKKAN
DATA MICROARRAY MENGGUNAKAN K-MEANS


Oleh
RADITYA RAKA ALIF AZIZI
1911016110001

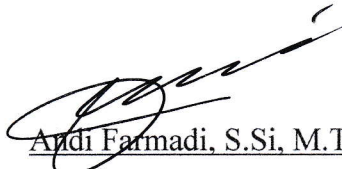
Telah dipertahankan di depan Dosen Penguji pada tanggal 17 Januari 2024.

Susunan Dosen Penguji:

Pembimbing I


Dosen Penguji I

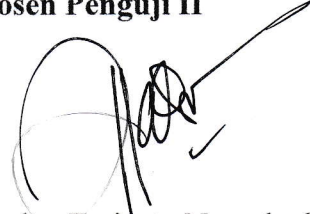

Dwi Kartini, S.Kom., M.Kom.
NIP. 198704212012122003


Andi Farmadi, S.Si, M.T
NIP. 197307252008011006

Pembimbing II

Dosen Penguji II


Irwan Budiman, S.T., M.Kom.
NIP. 197703252008121001


Dodon Turianto Nugrahadi, S.Kom, M.Eng.
NIP. 198001122009121002

Banjarbaru, 25 Januari 2024

Ketua 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, 4 Januari 2024



Raditya Raka Alif Azizi
NIM. 1911016110010

ABSTRAK

PERBANDINGAN PRINCIPAL COMPONENT ANALYSIS (PCA) DAN SPECTRAL CLUSTERING DALAM PENGELOMPOKKAN DATA MICROARRAY MENGGUNAKAN K-MEANS

(Oleh: Raditya Raka Alif Azizi; Pembimbing: Dwi Kartini, S.Kom., M.Kom.; Irwan Budiman, S.T., M.Kom.; 2023; 54 halaman)

Perbandingan reduksi dimensi Principal Component Analysis (PCA) dan Spectral Clustering yang mana kedua metode menggunakan nilai eigen dalam menentukan jumlah komponen yang akan direduksi. Data yang digunakan ialah ekspresi gen microarray dengan total 7129 data dan 34 fitur. Didapat pada kedua metode didapat hasil dimensi yang direduksi pada PCA yaitu 2 PC dan pada Spectral yaitu $k=2$ kemudian hasil reduksi dimensi tersebut dilakukan clustering menggunakan algoritma K-means. Hasil evaluasi pengelompokan menggunakan silhouette coefficient dan davies bouldin index (DBI) diperoleh nilai pada metode PCA K-means cluster optimal terdapat pada $k=2$ dengan nilai silhouette coefficient tertinggi 0,9480 dan DBI terendah 0,4065, sedangkan pada Spectral Clustering K-means cluster optimal terdapat pada pada $k = 2$ dengan nilai silhouette coefficient tertinggi 0,9311 dan DBI terendah 0,5536. Hasil penelitian menunjukkan PCA K-means memperoleh nilai evaluasi lebih tinggi dibandingkan Spectral Clustering K-means.

Kata kunci: *Microarray, Principal Analysis Component, Spectral Clustering, K-means.*

ABSTRACT

COMPARISON OF PRINCIPAL COMPONENT ANALYSIS (PCA) AND SPECTRAL CLUSTERING IN GROUPING MICROARRAY DATA USING K-MEANS

(By: Raditya Raka Alif Azizi; Supervisor: Dwi Kartini, S.Kom., M.Kom.; Irwan Budiman, S.T., M.Kom.; 2023; 54 pages)

Comparison of Principal Component Analysis (PCA) and Spectral Clustering dimension reduction where both methods use eigenvalues in determining the number of components to be reduced. The data used is microarray gene expression with a total of 7129 data and 34 features. The two methods obtained the results of the dimensions reduced in PCA, namely 2 PCs and in Spectral, namely $k = 2$, then the results of the dimensional reduction were carried out clustering using the K-means algorithm. The results of clustering evaluation using silhouette coefficient and davies bouldin index (DBI) obtained the value in the PCA K-means method the optimal cluster is at $k = 2$ with the highest silhouette coefficient value of 0.9480 and the lowest DBI of 0.4065, while in Spectral Clustering K-means the optimal cluster is at $k = 2$ with the highest silhouette coefficient value of 0.9311 and the lowest DBI of 0.5536. The results showed that PCA K-means obtained a higher evaluation value than Spectral Clustering K-means.

Keywords: Microarray, Principal Analysis Component, Spectral Clustering, K-means.

PRAKATA

Puji dan syukur kepada ALLAH SWT atas berkat rahmat-Nya penulis dapat menyelesaikan skripsi yang berjudul **PERBANDINGAN PRINCIPAL COMPONENT ANALYSIS (PCA) DAN SPECTRAL CLUSTERING DALAM PENGELOMPOKKAN DATA MICROARRAY MENGGUNAKAN K-MEANS** 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. Abah, mama, adik, 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 bapak Irwan Budiman, S.T., M.Kom. selaku dosen pembimbing utama dan pendamping pada pembuatan skripsi ini.
4. Bapak Andi Farmadi, S.Si, M.T. dan bapak Dodon Turianto Nugrahadi, S.Kom, M.Eng. 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. Falah Isnain, Muhammad Meftah Mafazy, Khairi Yahya, Muhammad Hanan Ababil, M. Fajar Aditya, Muhammad Thoriq Hidayat, 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.

7. Intelcore 2019 yang telah membantu dalam menjalankan perkuliahan bersama dari awal sampai akhir.
8. Seluruh mahasiswa Ilmu Komputer ULM yang telah membantu dalam keseharian di lingkungan kampus ULM Banjarbaru.
9. 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, 4 Januari 2024

Raditya Raka Alif Azizi

DAFTAR ISI

HALAMAN JUDUL	i
PERNYATAAN.....	iv
ABSTRAK	v
ABSTRACT.....	vi
PRAKATA.....	vii
DAFTAR ISI.....	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR.....	xiv
DAFTAR LAMPIRAN.....	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan.....	3
1.4 Manfaat Penelitian.....	4
1.5 Batasan Masalah.....	4
BAB II TINJAUAN PUSTAKA.....	5
Literatur Terdahulu.....	5
2.1 Microarray.....	6
2.2 Data mining.....	7
2.4 Z-Score.....	8
2.5 Principal Component Analysis.....	8
2.6 Spectral Clustering.....	9
2.7 K-Means.....	10
2.8 Silhoutte Coefficient	11
2.9 Davies Bouldin Index (DBI)	12
Keaslian Penelitian	13
BAB III METODE PENELITIAN.....	15

3.1	Alat Penelitian	15
3.2	Bahan Penelitian	15
3.3	Prosedur Penelitian.....	16
BAB IV HASIL DAN PEMBAHASAN		18
4.1	Hasil	18
4.1.1	Pengumpulan Data	18
4.1.2	Preprocessing.....	19
4.1.3	Modeling Data Mining	21
4.1.4	Evaluasi	37
4.2	Pembahasan.....	48
BAB V PENUTUP.....		55
5.1	Kesimpulan.....	55
5.2	Saran.....	55
LAMPIRAN.....		59