



SKRIPSI

IMPLEMENTASI METODE AFFINITY PROPAGATION (AP) DENGAN PRINCIPAL COMPONENT ANALYSIS (PCA) DALAM CLUSTERING DATA PASIEN HIPERTENSI

ANISSA ANDIAR BHALQIS
NPM 21083010038

DOSEN PEMBIMBING

Kartika Maulida Hindrayani, S.Kom., M.Kom.
Dr. Ir. Mohammad Idhom, S.P., S.Kom., M.T.

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SAINS DATA
SURABAYA
2026**

LEMBAR PENGESAHAN

IMPLEMENTASI METODE AFFINITY PROPAGATION (AP) DENGAN PRINCIPAL COMPONENT ANALYSIS (PCA) DALAM CLUSTERING DATA PASIEN HIPERTENSI

Oleh:

ANISSA ANDIAR BHALQIS
NPM. 21083010038

Telah dipertahankan di hadapan dan diterima oleh Tim Penguji Sidang Skripsi Program Studi Sains Data Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur pada Tanggal 26 Februari 2026:

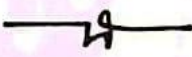
Menyetujui,

Kartika Maulida Hindrayani, S.Kom., M.Kom.
NIP. 19920909 202203 2 009


.....

(Pembimbing I)

Dr. Ir. Mohammad Idhom, S.P., S.Kom., M.T.
NIP. 19830310 202121 1 006


.....

(Pembimbing II)

Trimono, S.Si., M.Si.
NIP. 19950908 202203 1 003


.....

(Ketua Penguji)

Alfan Rizaldy Pratama, S.Tr.T., M.Tr.Kom.
NIP. 19990606 202406 1 001


.....

(Penguji I)

Mengetahui,

Dekan Fakultas Ilmu Komputer



Prof. Dr. Ir. Novirina Hendrasarie, MT.
NIP. 19681126 199403 2 001

LEMBAR PERSETUJUAN

**IMPLEMENTASI METODE AFFINITY PROPAGATION (AP) DENGAN
PRINCIPAL COMPONENT ANALYSIS (PCA) DALAM CLUSTERING
DATA PASIEN HIPERTENSI**

Oleh:
ANISSA ANDIAR BHALQIS
NPM. 21083010038

Telah disetujui untuk mengikuti Ujian Skripsi

Menyetujui,

**Koordinator Program Studi Sains Data
Fakultas Ilmu Komputer**

Dr. Eng. Ir. Dwi Arman Prasetya, S.T., M.T., IPU., Asean Eng.
NIP. 19801205 200501 1 002

SURAT PERNYATAAN BEBAS PLAGIASI

Saya yang bertanda tangan di bawah ini:

Nama Mahasiswa : Anissa Andiar Bhalqis
NPM : 21083010038
Program : Sarjana (S1)
Program Studi : Sains Data
Fakultas : Fakultas Ilmu Komputer

Menyatakan bahwa dalam dokumen ilmiah Skripsi ini tidak terdapat bagian dari karya ilmiah lain yang telah diajukan untuk memperoleh gelar akademik di suatu lembaga Pendidikan Tinggi, dan juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang/lembaga lain, kecuali yang secara tertulis disitasi dalam dokumen ini dan disebutkan secara lengkap dalam daftar pustaka.

Dan saya menyatakan bahwa dokumen ilmiah ini bebas dari unsur-unsur plagiasi. Apabila di kemudian hari ditemukan indikasi plagiat pada Skripsi ini, saya bersedia menerima sanksi sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sesungguhnya tanpa ada paksaan dari siapapun juga dan untuk dipergunakan sebagaimana mestinya.

Surabaya, 5 Maret 2026
Yang Membuat Pernyataan,



ANISSA ANDIAR BHALQIS
NPM. 21083010038

ABSTRAK

Nama Mahasiswa / NPM : Anissa Andiar Bhalqis / 21083010038
Judul Skripsi : Implementasi Metode Affinity Propagation (AP) dengan Principal Component Analysis (PCA) dalam Clustering Data Pasien Hipertensi
Dosen Pembimbing : 1. Kartika Maulida Hindrayani, S.Kom., M.Kom.
2. Dr. Ir. Mohammad Idhom, S.P., S.Kom., M.T.

Hipertensi adalah penyakit kronis yang ditandai oleh peningkatan tekanan darah dalam pembuluh arteri hingga melampaui nilai normal yang seharusnya. Data Profil Kesehatan Surabaya tahun 2019–2021 menunjukkan peningkatan prevalensi hipertensi dari 70,9% menjadi 84,9%, menegaskan bahwa kondisi ini menjadi masalah kesehatan serius karena berdampak pada penurunan kualitas hidup serta peningkatan beban layanan kesehatan. Data Kemenkes RI (2024) mencatat beban biaya BPJS akibat hipertensi mencapai Rp22,8 triliun dengan tingkat skrining rendah (3 dari 10 orang), sementara WHO (2023) menunjukkan hanya 15% pengobatan dan 4% kontrol tekanan darah. Kondisi ini menunjukkan bahwa pengelolaan hipertensi belum optimal dan masih bersifat seragam, sehingga berpotensi menyebabkan intervensi yang kurang tepat sasaran serta pemborosan sumber daya kesehatan. Oleh karena itu, untuk mendukung pendekatan penanganan yang lebih efisien, pendekatan berbasis data diperlukan untuk mengelompokkan pasien berdasarkan faktor risiko. Affinity Propagation adalah algoritma klasterisasi yang otomatis yang menentukan pusat klaster tanpa menetapkan jumlah klaster awal. Tujuan dari penelitian ini adalah untuk mengelompokkan pasien dengan hipertensi. AP bekerja melalui pertukaran pesan *responsibility* dan *availability* antar data untuk menentukan pusat klaster berdasarkan kemiripan. Untuk meningkatkan kualitas pengelompokan, diterapkan PCA sebagai tahap reduksi dimensi guna mengurangi redundansi dan *noise* data, yaitu variabel dengan variasi kecil yang sering dianggap kurang penting. Hasil reduksi menunjukkan empat variabel awal diringkas menjadi dua komponen utama. Dengan parameter AP berupa *preference* –500, *damping* 0,9, dan maksimum iterasi 200, diperoleh 7 klaster dengan *Silhouette Coefficient* sebesar 0,39 yang mengindikasikan kualitas klasterisasi tergolong sedang, sehingga pendekatan kombinasi PCA dan *Affinity Propagation* berpotensi mendukung pengambilan keputusan medis yang lebih terarah dalam pengelolaan hipertensi.

Kata Kunci: Hipertensi, *Clustering*, *Affinity Propagation*, *Principal Component Analysis*, *Silhouette Coefficient*

ABSTRACT

Student Name / NPM : Anissa Andiar Bhalqis / 21083010038
Undergraduate thesis title : *Implementation of Affinity Propagation (AP) Method with Principal Component Analysis (PCA) in Clustering Hypertension Patient Data*
Advisors : 1. Kartika Maulida Hindrayani, S.Kom., M.Kom.
2. Dr. Ir. Mohammad Idhom, S.P., S.Kom., M.T.

Hypertension is a chronic disease characterized by an increase in blood pressure in the arteries beyond normal levels. Data from the Surabaya Health Profile for 2019–2021 shows an increase in the prevalence of hypertension from 70.9% to 84.9%, confirming that this condition is a serious health problem because it reduces quality of life and increases the burden on health services. Data from the Indonesian Ministry of Health (2024) shows that the cost burden on BPJS due to hypertension reached IDR 22.8 trillion with a low screening rate (3 out of 10 people), while the WHO (2023) shows that only 15% of patients receive treatment and 4% have their blood pressure controlled. This condition indicates that hypertension management is not yet optimal and remains uniform, potentially leading to interventions that are not targeted and a waste of health resources. Therefore, to support a more efficient treatment approach, a data-driven approach is needed to group patients based on risk factors. Affinity Propagation is an automatic clustering algorithm that determines cluster centers without setting the initial number of clusters. The purpose of this study is to group patients with hypertension. AP works through the exchange of responsibility and availability messages between data to determine cluster centers based on similarity. To improve the quality of grouping, PCA is applied as a dimension reduction stage to reduce data redundancy and noise, namely variables with small variations that are often considered less important. The reduction results show that the four initial variables are summarized into two main components. With AP parameters of preference –500, damping 0.9, and maximum iterations of 200, seven clusters were obtained with a Silhouette Coefficient of 0.39, indicating moderate clustering quality. Thus, the combination of PCA and Affinity Propagation approaches has the potential to support more targeted medical decision-making in the management of hypertension.

Keywords: *Hypertension, Clustering, Affinity Propagation, Principal Component Analysis, Silhouette Coefficient*

KATA PENGANTAR

Puji syukur kehadirat Allah SWT atas segala rahmat, hidayah dan karunia-Nya kepada penulis sehingga proposal skripsi dengan judul **“Implementasi Metode Affinity Propagation (AP) Dengan Principal Component Analysis (PCA) Dalam Clustering Data Pasien Hipertensi”** dapat terselesaikan dengan baik.

Penulis mengucapkan terima kasih kepada Ibu Kartika Maulid Hindrayani, S.Kom., M.Kom. selaku Dosen Pembimbing I dan Bapak Dr. Ir. Mohammad Idhom, S.P., S.Kom., M.T. selaku Dosen Pembimbing 2 yang telah meluangkan waktunya untuk memberikan bimbingan, nasehat serta motivasi kepada penulis. Selain itu, selama penyusunan proposal skripsi penulis juga banyak menerima bantuan dari berbagai pihak. Untuk itu penulis mengucapkan terima kasih kepada:

1. Ibu Prof. Dr. Ir. Novirina Hendrasarie, M.T selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
2. Bapak Dr. Eng. Ir. Dwi Arman Prasetya., ST., MT., IPU., Asean., Eng. selaku Koordinator Program Studi Sains Data Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur.
3. Dosen-dosen Program Studi Sains Data yang telah membimbing dan memberikan ilmu yang bermanfaat bagi penulis selama masa perkuliahan
4. Allah SWT, Tuhan Yang Maha Esa, atas segala rahmat, hidayah, kesehatan, dan kemudahan yang dilimpahkan kepada penulis sehingga skripsi ini dapat terselesaikan dengan baik. Tanpa izin dan pertolongan-Nya, seluruh proses ini tidak akan berjalan dengan lancar.
5. Kepada kedua orang tua penulis, terima kasih atas segala doa, pengorbanan, dukungan moral maupun finansial, serta kepercayaan yang diberikan sehingga penulis dapat menyelesaikan pendidikan ini. Semoga Bapak dan Ibu senantiasa diberikan kesehatan dan kebahagiaan.
6. Adik Laki-Laki dan Keluarga besar penulis penulis yang memberikan dukungan, doa, dan semangat sehingga penulis bisa berada di titik ini.

7. Kepada sahabat sekaligus teman seperjuangan penulis Wenny, Mareta, Zulfa, Lisa, Enzel, Elina, Vera terima kasih telah menjadi teman tumbuh bersama disegala kondisi yang terkadang tidak terduga. Terima kasih selalu ada di sisi penulis dari semester awal hingga detik ini.
8. Kepada sahabat penulis sejak di bangku sekolah, Kwin dan Dinda, penulis mengucapkan terima kasih atas kesediaan dan kesabaran dalam mendengarkan setiap keluh kesah selama proses penyusunan skripsi ini, serta atas dukungan, bantuan, dan doa yang senantiasa diberikan.
9. Untuk seseorang yang selalu hadir dan setia memberikan dukungan, semangat, serta doa dalam setiap langkah penulis selama penyusunan skripsi ini, penulis menyampaikan terima kasih yang tulus dan mendalam.
10. Rekan-rekan mahasiswa Prodi Sains Data Universitas Pembangunan Nasional “Veteran” Jawa Timur terima kasih atas dukungan dan kerjasamanya selama menepuh Pendidikan serta penyelesaian penyusunan skripsi ini.
11. Yang terakhir, penulis ingin menyampaikan terima kasih kepada diri sendiri atas segala perjuangan, ketekunan, dan kesabaran dalam menghadapi setiap proses yang tidak selalu mudah. Terima kasih karena telah mampu bertahan, bangkit, dan terus melangkah dalam berbagai keadaan, serta tidak menyerah meskipun dihadapkan pada tantangan dan hambatan selama penyusunan skripsi ini. Terima kasih atas kerja keras, komitmen, dan keberanian untuk terus berusaha hingga akhirnya skripsi ini dapat terselesaikan dengan baik.

Penulis menyadari bahwa di dalam penyusunan skripsi ini banyak terdapat kekurangan. Untuk itu kritik dan saran yang membangun dari semua pihak sangat diharapkan demi kesempurnaan penulisan skripsi ini. Akhirnya, dengan segala keterbatasan yang penulis miliki semoga laporan ini dapat bermanfaat bagi semua pihak umumnya dan penulis pada khususnya.

Surabaya, Maret 2026

Penulis

DAFTAR ISI

HALAMAN SAMPUL.....	i
LEMBAR PENGESAHAN	ii
LEMBAR PERSETUJUAN	iii
SURAT PERNYATAAN BEBAS PLAGIASI.....	iv
ABSTRAK	v
ABSTRACT	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	ix
DAFTAR GAMBAR.....	xii
DAFTAR TABEL	xiii
DAFTAR LAMPIRAN	xiv
DAFTAR NOTASI.....	xv
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	5
1.3. Batasan Masalah.....	6
1.4. Tujuan Penelitian	6
1.5. Manfaat Penelitian	6
BAB II TINJAUAN PUSTAKA.....	9
2.1. Penelitian Terdahulu	9
2.2. Kerangka Teori.....	14
2.2.1. Hipertensi	15
2.2.2. <i>Clustering</i>	15
2.2.3. Standarisasi Data	16
2.2.4. Uji Asumsi PCA.....	18
2.2.5. <i>Principal Component Analysis (PCA)</i>	20
2.2.6. <i>Affinity Propagation (AP)</i>	24
2.2.7. <i>Silhouette Coefficient</i>	28

2.2.8.	<i>Graphical User Interface (GUI)</i>	31
2.2.9.	<i>Streamlit</i>	31
BAB III METODOLOGI PENELITIAN.....		33
3.1.	Variabel Penelitian dan Sumber Data.....	33
3.2.	Langkah Analisis	34
3.2.1.	Pengumpulan Data.....	35
3.2.2.	Persiapan Data	35
3.2.3.	Uji Asumsi PCA	36
3.2.4.	<i>Principal Component Analysis (PCA)</i>	36
3.2.5.	<i>Affinity Propagation (AP)</i>	36
3.2.6.	Evaluasi Model	38
3.2.7.	Hasil <i>Clustering</i>	38
3.2.8.	Implementasi GUI	39
3.3.	Desain Sistem	39
BAB IV HASIL DAN PEMBAHASAN.....		45
4.1.	Hasil Penelitian.....	45
4.1.1.	Pengumpulan data	45
4.1.2.	Persiapan Data	47
4.1.3.	Uji Asumsi PCA	58
4.1.4.	<i>Principal Component Analysis (PCA)</i>	61
4.1.5.	Pemodelan <i>Affinity Propagation (AP)</i>	72
4.1.6.	Evaluasi Model	80
4.1.7.	Hasil <i>Clustering</i>	83
4.2.	Tampilan Antar Muka (<i>Graphical User Interface</i>)	90
4.2.1.	Halaman Utama	90
4.2.2.	Halaman <i>Input Data</i>	91
4.2.3.	Halaman <i>Preprocessing</i>	92
4.2.4.	Halaman Standarisasi	93
4.2.5.	Halaman PCA	93
4.2.6.	Halaman AP.....	94
4.2.7.	Halaman Visualisasi	94

BAB V PENUTUP	97
5.1. Kesimpulan	97
5.2. Saran Pengembangan	98
DAFTAR PUSTAKA	99
LAMPIRAN.....	105

DAFTAR GAMBAR

Gambar 2.1. Ilustrasi <i>Similarity</i>	26
Gambar 3.1. Diagram Alir Penelitian.....	34
Gambar 3.2. Diagram Alir Desain Sistem.....	40
Gambar 3.3. <i>Wireframe</i> Halaman Utama.....	42
Gambar 3.4. <i>Wireframe</i> Halaman Analisis.....	42
Gambar 4.1. Hasil Cek Data Duplikat.....	50
Gambar 4.2. Boxplot <i>Outlier</i>	51
Gambar 4.3. Boxplot <i>Winsorizing</i>	55
Gambar 4.4. Hasil Uji KMO.....	59
Gambar 4.5. Hasil Uji Bartlett.....	61
Gambar 4.6. <i>Scree Plot</i>	68
Gambar 4.7. Visualisasi Distribusi <i>Cluster</i>	85
Gambar 4.8. Karakteristik Tiap <i>Cluster</i>	86
Gambar 4.9. Visualisasi <i>Cluster</i> AP+PCA.....	87
Gambar 4.10. Visualisasi <i>Cluster</i> DBSCAN.....	88
Gambar 4.11. Tampilan Halaman Utama.....	90
Gambar 4.12. Halaman <i>Input Data</i>	91
Gambar 4.13. Tampilan Halaman <i>Input Data</i>	91
Gambar 4.14. Tampilan Halaman <i>Preprocessing</i>	92
Gambar 4.15. Tampilan Halaman Standarisasi.....	93
Gambar 4.16. Tampilan Halaman PCA.....	93
Gambar 4.17. Tampilan Halaman AP.....	94
Gambar 4.18. Tampilan Halaman Visualisasi.....	95

DAFTAR TABEL

Tabel 2.1. Penelitian Terdahulu.....	9
Tabel 2.2. Interpretasi Nilai <i>Silhouette Coefficient</i>	30
Tabel 3.1. Variabel Penelitian	33
Tabel 3.2. Struktur Data	33
Tabel 4.1. Data Pasien Hipertensi	45
Tabel 4.2. Analisis Statistik Deskriptif	46
Tabel 4.3. Hasil Pengecekan <i>Missing Value</i>	48
Tabel 4.4. Hasil Penanganan <i>Missing Value</i>	49
Tabel 4.5. Hasil Deteksi <i>Outlier</i>	53
Tabel 4.6. Hasil Standarisasi Data.....	57
Tabel 4.7. Hasil Varian-Kovarians	63
Tabel 4.8. Hasil Perhitungan Varians-Kovarians	64
Tabel 4.9. Hasil Nilai dan Vektor Eigen	65
Tabel 4.10. Hasil Perhitungan Nilai dan Vektor Eigen.....	67
Tabel 4.11. Hasil Kumulatif Varians.....	69
Tabel 4.12. Hasil Perhitungan Kumulatif Varians	70
Tabel 4.13. Hasil Proyeksi Data Baru	71
Tabel 4.14. Hasil Perhitungan Proyeksi Data.....	72
Tabel 4.15. Hasil Menghitung Nilai <i>Similarity</i>	74
Tabel 4.16. Hasil Percobaan <i>Preference</i>	75
Tabel 4.17. Hasil Perhitungan Iterasi 1	78
Tabel 4.18. Hasil Pemodelan <i>Affinity Propagation</i>	80
Tabel 4.19. Evaluasi <i>Silhouette Coefficient</i> AP.....	81
Tabel 4.20. Rasio Tiap <i>Cluster</i> AP+PCA.....	81
Tabel 4.21. Rasio Tiap <i>Cluster</i> DBSCAN.....	82
Tabel 4.22. Hasil Pengelompokan Data Pasien.....	83
Tabel 4.23. Distribusi Data.....	84
Tabel 4.24. Karakteristik Data Tiap <i>Cluster</i>	85

DAFTAR LAMPIRAN

Lampiran 1. <i>Source Code</i> AP dengan PCA	105
Lampiran 2. <i>Source Code</i> GUI.....	106

DAFTAR NOTASI

z	:	Data yang distandarisasi
$\bar{x}$	:	Rata-rata dari data
s^2	:	Varian
s	:	Standar Deviasi
x_i	:	Nilai dari setiap data ke- i
n	:	Jumlah data
$Var(X_p)$	:	Sebaran data variabel terhadap rata-ratanya
$X_{p,i}$	:	Nilai ke- i dari variabel X_p
$\bar{X}_p$	:	Rata-rata dari variabel X_p
$\bar{X}_{p+1}$	:	Rata-rata dari variabel X_{p+1}
$\mathbf{A}$	:	matriks persegi
v	:	vektor
λ	:	skalar
$\mathbf{I}$	:	Matriks Identitas
e_i	:	Eigen vektor ke- i
y	:	Dataset baru setelah dilakukan PCA
$\mathbf{Z}$	:	Matriks dataset awal hasil dari standarisasi data
$\mathbf{W}$	:	Matriks eigen vektor
$S_{(i,j)}$	:	Nilai kesamaan (<i>similarity</i>) antara titik data i dan j
x_{ik}	:	Nilai pada dimensi ke- k untuk titik data ke- i
x_{jk}	:	Nilai pada dimensi ke- k untuk titik data ke- j
n	:	Jumlah total dimensi atau variabel dalam data
$r(i, k)$	:	Nilai <i>responsibility</i> yang menunjukkan seberapa cocok X_k sebagai pusat <i>cluster</i> untuk X_i
$s(i, k)$	:	Nilai <i>similarity</i> antara X_i dan X_k
$\max_{k' \neq k} \{a(i, k') + s(i, k')\}$	:	Nilai maksimum gabungan <i>availability</i> dan <i>similarity</i> untuk titik data selain k

$r(j, j)$	:	<i>Responsibility</i> untuk X_j sebagai pusat <i>cluster</i> bagi dirinya sendiri
$s(k, k)$	:	Nilai ksesamaan atau <i>self-similarity</i> titik data X_k
$\max\{s(i, k')\}$	:	Nilai maksimum kesamaan antara X_i dan kemungkinan lain $X_{k'}$
$a(i, k)$	:	<i>Availability</i> yang mengukur sejauh mana kandidat X_k dapat dipilih sebagai pusat <i>cluster</i> bagi titik data X_i
$\min\{0, \dots\}$	:	Nilai <i>availability</i> dibatasi agar tidak lebih besar dari 0
$r(k, k)$	:	<i>Responsibility</i> yang menunjukkan bahwa X_k sebagai pusat <i>cluster</i> untuk dirinya sendiri
$\sum_{i \in \{i, k\}} \max\{0, r(i', k)\}$	:	Menjumlahkan <i>responsibility</i> dari $X_{i'}$ untuk kandidat pusat <i>cluster</i> X_k
$a(k, k)$	:	Menunjukkan <i>availability</i> dari titik data X_k untuk menjadi pusat <i>cluster</i> bagi diri sendiri
$c(i, k)$	:	Kombinasi dari <i>availability</i> dan <i>responsibility</i> untuk titik data X_i terhadap kandidat pusat <i>cluster</i> X_k
$k \leftarrow \arg \max$	:	Menentukan k pusat <i>cluster</i> dengan nilai maksimum dari jumlah <i>responsibility</i> dan <i>availability</i>
$S(i)$	:	Nilai <i>Silhouette Coeficient</i> untuk objek ke- i
$a(i)$	:	Rata-rata jarak antara data i dengan semua data lain dalam <i>cluster</i> yang sama
$b(i)$	:	Rata-rata jarak antara data i dengan semua data dalam <i>cluster</i> terdekat yang berbeda
$\max(a(i), b(i))$	:	Mengambil nilai maksimum antara $a(i)$ dan $b(i)$
e_q	:	<i>Exemplar</i> dari klaster C_q
c_X	:	Pusat keseluruhan data
n_q	:	Jumlah anggota dalam klaster C_q
x	:	Setiap titik data dalam klaster C_q
k	:	Jumlah klaster yang terbentuk
C_i	:	Jumlah titik data dalam klaster C_i

x_j	:	Titik data dalam klaster C_i
E_i	:	<i>Exemplar</i> untuk klaster C_i
$d(x_j, E_i)$	:	Jarak Euclidean antara titik data x_j dan <i>exemplar</i> E_i
$d(i, j)$	:	Jarak antar klaster i dan j
S_i	:	Dispersi dalam klaster i
S_j	:	Dispersi dalam klaster j
N	:	Jumlah total klaster
$\max_{j \neq i} R_{ij}$	:	Nilai maksimum dari rasio R_{ij} untuk setiap klaster i
p	:	Jumlah variabel
r_{ij}	:	Koefisien korelasi antara variabel i dan j
p_{ij}	:	Koefisien korelasi parsial antara variabel i dan j
a_{ij}	:	Elemen matriks <i>inverse</i>
χ^2	:	Nilai statistik uji Bartlett
$ R $	:	Determinan matriks korelasi
$\ln$	:	Logaritma natural
df	:	Derajat bebas