

PERBANDINGAN TEKNIK REDUKSI DIMENSI ANTARA
ALGORITMA *PRINCIPAL COMPONENT ANALYSIS* DENGAN
FUZZY ASSOCIATION RULE TERHADAP HASIL
PENGLASTERAN

*Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Studi Strata-1 Pada
Jurusan Teknik Informatika Reguler Fakultas Ilmu Komputer*



Oleh :

RADEN RORO AYU LARASWATI
NIM : 09021181520003

**JURUSAN TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA**

2020

LEMBAR PENGESAHAN TUGAS AKHIR

Perbandingan Teknik Reduksi Dimensi Antara Algoritma *Principal Component Analysis* dengan *Fuzzy Association Rule* Terhadap Hasil Pengklasteran

Oleh :

RADEN RORO AYU LARASWATI
NIM : 09021181520003

Indralaya, 28 November 2020

Pembimbing I,



Ir. M. Ihsan Jambak, M.Sc.
NIP. 196804052013081201

Pembimbing II,



Desty Rodiah, M.T.
NIP. 1671016112890005

Mengetahui,

Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

TANDA LULUS SIDANG TUGAS AKHIR

Pada hari Rabu, 18 November 2020 telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Raden Roro Ayu Laraswati
NIM : 09021181520003
Judul : Perbandingan Teknik Reduksi Dimensi Antara Algoritma *Principal Component Analysis* dengan *Fuzzy Association Rule* Terhadap Hasil Pengklasteran

1. Pembimbing I

Ir. M. Ihsan Jambak, M.Sc.
NIP. 196804052013081201



2. Pembimbing II

Desty Rodiah, M.T.
NIP. 1671016112890005



3. Penguji I

Novi Yusliani, M.T.
NIP. 198211082012122001



4. Penguji II

Danny Matthew Saputra, M.Sc.
NIP. 198505102015041002



Mengetahui,
Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Raden Roro Ayu Laraswati

NIM : 09021181520003

Program Studi : Teknik Informatika

Judul Skripsi : Perbandingan Teknik Reduksi Dimensi Antara Algoritma *Principal Component Analysis* dengan *Fuzzy Association Rule* Terhadap Hasil Pengklasteran

Hasil Pengecekan *Software iThenticate/Turnitin* : 9%

Menyatakan bahwa Laporan Tugas Akhir saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan projek ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.



MOTTO DAN PERSEMBAHAN

Motto :

- ❖ It's Okay To Stop For A While, And Then Continue With A Big Step
- ❖ If You Feel Down, Comfort Yourself by Talk To Your Reflection

Kupersembahkan karya tulis ini kepada :

- Allah SWT
- Diriku Sendiri
- Ibu dan Alm. Ayahku tercinta
- Adikku tersayang
- Keluarga besarku
- Sahabat dan teman seperjuanganku
- Jurusan Teknik Informatika
- Fakultas Ilmu Komputer
- Universitas Sriwijaya

KATA PENGANTAR



Puji syukur kepada Allah SWT atas berkat dan rahmat-Nya yang telah diberikan kepada Penulis sehingga Penulis dapat menyelesaikan Tugas Akhir ini dengan baik. Tugas akhir ini disusun untuk memenuhi salah satu syarat menyelesaikan pendidikan program Strata-1 pada Fakultas Ilmu Komputer Program Studi Teknik Informatika di Universitas Sriwijaya.

Dalam menyelesaikan Tugas Akhir ini banyak sekali pihak yang telah memberikan bantuan dan dukungan baik secara langsung maupun secara tidak langsung. Penulis ingin menyampaikan rasa terima kasih kepada:

1. Ibuku Murdiana, Ayahku Alm. R. Iwan Iriawan, Adikku Raden Dimas Putra Iriawan, dan seluruh keluarga besarku yang selalu mendokan serta memberikan dukungan kepadaku.
2. Bapak Jaidan Jauhari, S.Pd., M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
3. Ibu Alvi Syahrini Utami, M.Kom. selaku Ketua Jurusan Teknik Informatika.
4. Bapak Drs. Megah Mulya, M.T. selaku dosen pembimbing akademik yang telah membimbing, mengarahkan dan memberikan motivasi Penulis dalam proses perkuliahan.
5. Bapak Ir. M. Ihsan Jambak, M.Sc. selaku dosen pembimbing I dan Ibu Desty Rodiah, M.T. selaku dosen pembimbing II yang telah membimbing, mengarahkan dan memberikan motivasi Penulis dalam proses perkuliahan dan pengerjaan Tugas Akhir.

6. Ibu Novi Yusliani, M.T. selaku penguji I dan Bapak Danny Matthew Saputra., M.Sc. selaku dosen penguji II yang telah memberikan masukan dan dorongan dalam proses pengerjaan Tugas Akhir.
7. Seluruh dosen Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
8. Mbak Winda, Mbak Wiwin, Kak Hafez, Kak Ricy serta seluruh staf tata usaha yang telah membantu dalam kelancaran proses administrasi dan akademik selama masa perkuliahan.
9. Sahabat Astaghfirullahku, Divin, Okta, Imas, Nadiya dan Elin yang telah menemani dengan canda, tawa, dan air mata selama masa perkuliahan, semoga kita tetap bersahabat sampai tua bersama.
10. Kedua sahabatku Ayum dan Diah Ayu yang kurindukan, yang telah menjadi tempat berbagi cerita dari masa SMA sampai sekarang, semoga kita segera bertemu kembali.
11. Sahabatku Rista, Adik Acehku Rossi yang selalu cerewet ngajak wisuda bareng. Kakak, Mbak, Adik, dan Teman-teman BPH dan seper-Ki-anku di UKM U-Read 2018/2019.
12. Teman-teman seperjuanganku sampai akhir, Anggik, Syazili, Nopren, Azed yang telah membantu dan bergadang bersama untuk menyelesaikan Tugas Akhir ini, serta teman-teman seangkatan IF Reguler 2015 yang telah berjuang bersama dalam menimba ilmu di Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Penulis menyadari dalam penyusunan Tugas Akhir ini masih banyak terdapat kekurangan disebabkan keterbatasan pengetahuan dan pengalaman, oleh karena itu kritik dan saran yang membangun sangat diharapkan untuk kemajuan penelitian selanjutnya.

Akhir kata semoga Tugas Akhir ini dapat berguna dan bermanfaat bagi kita semua.

Indralaya, 28 November 2020



Raden Roro Ayu Laraswati

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
LEMBAR PENGESAHAN TUGAS AKHIR	ii
TANDA LULUS SIDANG TUGAS AKHIR	iii
HALAMAN PERNYATAAN	iv
MOTTO DAN PERSEMBAHAN	v
ABSTRACT	vi
ABSTRAK	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	xi
DAFTAR TABEL	xiv
DAFTAR GAMBAR	xvi
DAFTAR LAMPIRAN	xviii
BAB I PENDAHULUAN	
1.1 Pendahuluan	I-1
1.2 Latar Belakang Masalah	I-1
1.3 Rumusan Masalah	I-4
1.4 Tujuan Penelitian	I-4
1.5 Manfaat Penelitian	I-5
1.6 Batasan Masalah	I-5
1.7 Kesimpulan	I-5
BAB II KAJIAN LITERATUR	
2.1 Pendahuluan	II-1
2.2 Konversi Data	II-1
2.3 Pengklasteran <i>k-Means</i>	II-3
2.4 Reduksi Dimensi	II-6
2.4.1 <i>Principal Component Analysis (PCA)</i>	II-7
2.4.2 <i>Fuzzy Association Rule (FAR)</i>	II-9
2.5 <i>Davies Boundin Index (DBI)</i>	II-11
2.6 <i>Rational Unified Process (RUP)</i>	II-13
2.7 Penelitian Lain Yang Relevan.....	II-15
2.8 Kesimpulan	II-18
BAB III METODOLOGI PENELITIAN	
3.1 Pendahuluan	III-1
3.2 Unit Penelitian	III-1
3.3 Pengumpulan Data	III-1
3.3.1 Jenis dan Sumber Data	III-1
3.3.2 Metode Pengumpulan Data	III-2

3.4 Tahapan Penelitian	III-2
3.4.1 Menetapkan Kerangka Kerja	III-2
3.4.1.1 Konversi Format Data	III-2
3.4.1.2 Nilai k Optimum pada <i>k-Means</i>	III-4
3.4.1.3 Reduksi Dimensi	III-4
3.4.1.4 Pengklasteran Hasil Reduksi Dimensi Dengan Jumlah Klaster Optimum	III-5
3.4.1.5 Evaluasi Pengklasteran	III-6
3.4.2 Menetapkan Kriteria Pengujian	III-7
3.4.3 Menetapkan Format Data Pengujian	III-7
3.4.4 Menentukan Alat yang Digunakan dalam Pelaksanaan Penelitian	III-8
3.4.5 Melakukan Tahapan Penelitian	III-9
3.4.6 Melakukan Analisa Hasil Pengujian dan Membuat Kesimpulan	III-10
3.5 Metode Pengembangan Perangkat Lunak	III-11
3.5.1 <i>Rational Unified Process</i> (RUP)	III-11
3.5.2 Fase Insepsi (Permulaan)	III-12
3.5.3 Fase Elaborasi (Perencanaan)	III-13
3.5.4 Fase Konstruksi (Pengembangan)	III-13
3.5.5 Fase Transisi	III-14
3.6 Manajemen Proyek Penelitian	III-14
3.7 Kesimpulan	III-23

BAB IV PENGEMBANGAN PERANGKAT LUNAK

4.1 Pendahuluan	IV-1
4.2 Fase Insepsi	IV-1
4.2.1 Pemodelan Bisnis	IV-2
4.2.2 Kebutuhan Sistem	IV-4
4.2.2.1 Fungsi Muat Data	IV-5
4.2.2.2 Fungsi Reduksi Dimensi	IV-7
4.2.2.3 Fungsi Pengklasteran	IV-13
4.2.3 Analisa dan Desain	IV-15
4.3 Fase Elaborasi	IV-17
4.3.1 Pemodelan Bisnis	IV-17
4.3.2 Kebutuhan Sistem	IV-23
4.3.3 Analisa Desain	IV-27
4.4 Fase Konstruksi	IV-32
4.4.1 Pemodelan Bisnis	IV-32
4.4.2 Kebutuhan Sistem	IV-34
4.4.3 Implementasi	IV-34
4.4.4 Implementasi Kelas	IV-34
4.4.5 Rencana Pengujian	IV-36
4.4.5.1 Rencana Pengujian <i>Use Case</i> Melakukan Konversi Format Data Teks Menjadi Numerik.....	IV-36
4.4.5.2 Rencana Pengujian <i>Use Case</i> Mengklaster Data	

Menggunakan <i>k-Means</i>	IV-37
4.4.5.3 Rencana Pengujian <i>Use Case</i> Mereduksi Dimensi Data dengan <i>Fuzzy Association</i> <i>Rule</i> dan Pengklasteran dengan <i>k-Means</i>	IV-37
4.4.5.4 Rencana Pengujian <i>Use Case</i> Mereduksi Dengan PCA dan Pengklasteran dengan <i>k-Means</i>	IV-38
4.5 Fase Transisi	IV-39
4.5.1 Pemodelan Bisnis	IV-39
4.5.2 Kebutuhan Sistem	IV-40
4.5.3 Pengujian.....	IV-40
4.6 Kesimpulan	IV-45

BAB V HASIL DAN ANALISIS PENELITIAN

5.1 Pendahuluan	V-1
5.2 Data Hasil Uji Penelitian	V-1
5.2.1 Hasil Pengujian Penentuan Nilai K Optimum	V-1
5.2.2 Hasil Pengujian Pengklasteran <i>k-Means</i>	V-2
5.2.3 Hasil Pengujian Pengklasteran dengan Reduksi PCA	V-3
5.2.4 Hasil Pengklasteran dengan Reduksi <i>Fuzzy Association Rule</i>	V-4
5.2.4.1 Hasil Pengujian Nilai <i>Support</i> Optimum	V-5
5.2.4.2 Hasil Pengujian Pengklasteran dengan Reduksi <i>Fuzzy Association Rule</i>	V-6
5.3 Olah Data Hasil Uji Penelitian.....	V-7
5.3.1 Olah Data Hasil Pengujian K Optimum.....	V-7
5.3.2 Olah Data Hasil Perbandingan Nilai DBI.....	V-8
5.3.3 Olah Data Hasil Perbandingan Waktu Komputasi.....	V-15
5.3.4 Olah Data Hasil Perbandingan Jumlah Iterasi	V-19
5.4 Analisa Hasil Pengujian	V-25
5.4 Kesimpulan	V-28

BAB VI KESIMPULAN DAN SARAN

6.1 Pendahuluan	VI-1
6.2 Kesimpulan	VI-1
6.3 Saran	VI-2

DAFTAR PUSTAKA	xxi
LAMPIRAN	xxiii

DAFTAR TABEL

	Halaman
Tabel III-1 Rancangan Tabel Hasil Pengujian Penentuan Nilai k Optimum .	III-7
Tabel III-2 Tabel Evaluasi Klaster Internal Dengan PCA	III-8
Tabel III-3 Rancangan Tabel Evaluasi Klaster Internal Dengan <i>Fuzzy</i>	III-8
Tabel III-4 Rancangan Tabel Hasil Analisa Pengklasteran	III-11
Tabel III-5 Penjadwalan Penelitian dalam Bentuk <i>Work Breakdown Structure</i> (WBS)	III-15
Tabel IV-1. Definisi Aktor <i>Use Case</i>	IV-3
Tabel IV-2. Definisi <i>Use Case</i>	IV-3
Tabel IV-3. Kebutuhan Fungsional.....	IV-4
Tabel IV-4. Kebutuhan Non-Fungsional.....	IV-5
Tabel IV-5. Skenario Melakukan Muat Data.....	IV-6
Tabel IV-6. Skenario <i>Use Case</i> Mereduksi Dimensi Data Menggunakan PCA dan Pengklasteran <i>k-Means</i>	IV-8
Tabel IV-7. Skenario <i>Use Case</i> Mereduksi Dimensi Data Menggunakan <i>Fuzzy Association Rule</i> dan Pengklasteran <i>k-Means</i>	IV-11
Tabel IV-8. Skenario <i>Use Case</i> Melakukan Pengklasteran Menggunakan <i>k-Means</i>	IV-13
Tabel IV-9. Implementasi Kelas	IV-34
Tabel IV-10. Rencana Pengujian <i>Use Case</i> Melakukan Konversi Format Data Text Menjadi Numerik.....	IV-36
Tabel IV-11. Rencana Pengujian <i>Use Case</i> Melakukan Pengklasteran dengan <i>k-Means</i>	IV-37
Tabel IV-12. Rencana Pengujian <i>Use Case</i> Mereduksi dengan <i>Fuzzy Association Rule</i> dan Pengklasteran dengan <i>k-Means</i>	IV-38
Tabel IV-13. Rencana Pengujian <i>Use Case</i> Mereduksi dengan PCA dan Pengklasteran dengan <i>k-Means</i>	IV-38
Tabel IV-14. Pengujian Fitur Konversi Data.....	IV-41
Tabel IV-15. Pengujian Fitur Pengklasteran <i>k-Means</i>	IV-42
Tabel IV-16. Pengujian Fitur Pengklasteran <i>k-Means</i> dengan Reduksi dimensi PCA	IV-43
Tabel IV-17. Pengujian Fitur Pengklasteran <i>k-Means</i> dengan Reduksi dimensi <i>Fuzzy Association Rule</i>	IV-44
Tabel V-1. Hasil Pengujian Penentuan Nilai K Optimum.....	V-2
Tabel V-2. Hasil Pengujian Pengklasteran <i>k-Means</i>	V-2
Table V-3. Hasil Pengujian Pengklasteran dengan Reduksi PCA	V-4
Tabel V-4. Hasil Pengujian Penentuan Nilai <i>Support</i> Optimum	V-5
Tabel V-5. Hasil Pengujian Pengklasteran dengan Reduksi <i>Fuzzy Association Rule</i>	V-6
Tabel V-6. Deskripsi Nilai DBI	V-8
Tabel V-7. Hasil Uji Normalitas	V-9
Tabel V-8. Hasil Uji Kruskal-Wallis Perbandingan Nilai DBI	V-10

Tabel V-9. Hasil Tes Statistik Uji Kruskal-Wallis Perbandingan Nilai DBI.....	V-11
Tabel V-10. Hasil Uji Mann-Whitney Perbandingan Nilai DBI Pada Data Hasil Pengklasteran K-Means dan PCA+K-Means.....	V-11
Tabel V-11. Hasil Tes Statistik Uji Mann-Whitney Perbandingan Nilai DBI pada Data Hasil Pengklasteran <i>k-Means</i> dan PCA+ <i>k-Means</i>	V-12
Tabel V-12. Hasil Uji Mann-Whitney Perbandingan Nilai DBI Pada Data Hasil Pengklasteran PCA+K-Means dan FAR+K-Means.....	V-12
Tabel V-13. Hasil Tes Statistik Uji Mann-Whitney Perbandingan Nilai DBI pada Data Hasil Pengklasteran <i>k-Means</i> dan FAR+ <i>k-Means</i>	V-13
Tabel V-14. Hasil Uji Mann-Whitney Perbandingan Nilai DBI pada Data Hasil Pengklasteran PCA+ <i>k-Means</i> dan FAR+ <i>k-Means</i> ...	V-14
Tabel V-15. Hasil Tes Statistik Uji Mann-Whitney Perbandingan Nilai DBI pada Data Hasil Pengklasteran PCA+ <i>k-Means</i> dan FAR+ <i>k-Means</i>	V-14
Tabel V-16. Deskripsi Nilai Waktu Komputasi	V-15
Tabel V-17. Hasil Uji Normalitas Waktu Komputasi	V-16
Tabel V-18. Hasil Uji Homogenitas Waktu Komputasi	V-16
Tabel V-19. Hasil Uji ANOVA Waktu Komputasi	V-17
Tabel V-20. Hasil Uji Post Hoc Waktu Komputasi	V-18
Tabel V-21. Hasil Deskripsi Nilai Jumlah Iterasi	V-19
Tabel V-22. Hasil Uji Normalitas Jumlah Iterasi	V-20
Tabel V-23. Hasil Uji Kruskal-Wallis Perbandingan Nilai Jumlah Iterasi	V-21
Tabel V-24. Hasil Uji Kruskal-Wallis Perbandingan Nilai Jumlah Iterasi....	V-21
Tabel V-25. Hasil Uji Mann-Whitney Perbandingan Nilai Jumlah Iterasi pada Data Hasil Pengklasteran <i>k-Means</i> dan PCA+ <i>k-Means</i>	V-22
Tabel V-26. Hasil Tes Statistik Uji Mann-Whitney Perbandingan Nilai Jumlah IterasiI pada Data Hasil Pengklasteran <i>k-Means</i> dan PCA+ <i>k-Means</i>	V-23
Tabel V-27. Hasil Uji Mann-Whitney Perbandingan Nilai Jumlah Iterasi pada Data Hasil Pengklasteran <i>k-Means</i> dan FAR+ <i>k-Means</i>	V-23
Tabel V-28. Hasil Tes Statistik Uji Mann-Whitney Perbandingan Nilai Jumlah Iterasi pada Data Hasil Pengklasteran <i>k-Means</i> dan FAR+ <i>k-Means</i>	V-24
Tabel V-29. Hasil Uji Mann-Whitney Perbandingan Jumlah Iterasi pada Data Hasil Pengklasteran PCA+ <i>k-Means</i> dan FAR+ <i>k-Means</i> ...	V-24
Tabel V-30. Hasil Tes Statistik Uji Mann-Whitney Perbandingan Nilai Jumlah Iterasi pada Data Hasil Pengklasteran PCA+ <i>k-Means</i> dan FAR+ <i>k-Means</i>	V-25

DAFTAR GAMBAR

	Halaman
Gambar II-1. Nilai Keanggotaan Fuzzy Set Tipe-2	II-9
Gambar II-2. Arsitektur RUP	II-13
Gambar III-1. Tahap Pengujian Penelitian.....	III-10
Gambar III-2. Arsitektur RUP	III-12
Gambar III-3. Penjadwalan Penelitian Tahap Menentukan Ruang Lingkup dan Unit Penelitian.....	III-19
Gambar III-4. Penjadwalan Penelitian Tahap Menentukan Dasar Teori yang Berkaitan dengan Penelitian dan Menentukan Kriteria Pengujian	III-20
Gambar III-5. Penjadwalan Penelitian Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian pada Fase Insepsi ..	III-20
Gambar III-6. Penjadwalan Penelitian Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian pada Fase Elaborasi	III-21
Gambar III-7. Penjadwalan Penelitian Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian pada Fase Konstruksi	III-21
Gambar III-8. Penjadwalan Penelitian Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian pada Fase Transisi.	III-22
Gambar III-9. Penjadwalan Penelitian Tahap Melakukan Pengujian Penelitian, Penjadwalan Penelitian Tahap Melakukan Analisa Hasil Pengujian dan Membuat Kesimpulan.....	III-22
Gambar IV-1. Diagram <i>Use Case</i>	IV-2
Gambar IV-2. Diagram <i>Sequence</i> Melakukan Konversi Data	IV-19
Gambar IV-3. Diagram <i>Sequence</i> Melakukan Pengklasteran <i>k-Means</i>	IV-20
Gambar IV-4. Diagram <i>Sequence</i> Mereduksi Data dengan PCA dan Mengklaster Data Menggunakan <i>k-Means</i>	IV-21
Gambar IV-5. Diagram <i>Sequence</i> Mereduksi Data dengan <i>Fuzzy Association Rule</i> dan Mengklaster Data Menggunakan <i>k-Means</i>	IV-22
Gambar IV-6. Rancangan Antar Muka Perangkat Lunak	IV-23
Gambar IV-7. Rancangan Muat Data.....	IV-23
Gambar IV-8. Rancangan Fitur Input Jumlah Kluster	IV-24
Gambar IV-9. Rancangan Fitur Metode Pengklasteran.....	IV-24
Gambar IV-10. Rancangan Fitur Mulai Proses.....	IV-25
Gambar IV-11. Rancangan Fitur Waktu Komputasi, Jumlah Iterasi dan Nilai DBI.....	IV-26
Gambar IV-12. Rancangan Fitur Menampilkan Hasil Pengklasteran Dokumen Hasil Kluster.....	IV-26
Gambar IV-13. Diagram Aktivitas Muat Data.....	IV-28
Gambar IV-14. Diagram Aktivitas Mengklaster Data Menggunakan	

	<i>k-Means</i>	IV-29
Gambar IV-15. Diagram Aktivitas Mereduksi Data dengan PCA dan Mengklaster Data Menggunakan <i>k-Means</i>		IV-30
Gambar IV-16. Diagram Aktivitas Mereduksi Data dengan <i>Fuzzy</i> <i>Association Rule</i> dan Mengklaster Data Menggunakan <i>k-Means</i>		IV-31
Gambar IV-17. Diagram Kelas.....		IV-33
Gambar V-1. Kurva Hasil Penentuan K Optimum		V-8

DAFTAR LAMPIRAN

Lampiran I Kode Program

COMPARISON OF DIMENSIONAL REDUCTION USING *PRINCIPAL COMPONENT ANALYSIS* AND *FUZZY ASSOCIATION RULE* FOR CLUSTERING RESULTS

ABSTRACT

Clustering is the process of grouping data into several groups, where each member of the group has big similarities and disimilarities to other member of the group. In clustering, conventional algorithm works well in handling low-dimensional data, therefore to improve the quality of text clustering results, dimensional reduction technique is required. Dimensional reduction techniques are classified into 2 types, feature selection and feature extraction. This study will compare the application of the *Principal Component Analysis* (PCA) and *Fuzzy Association Rule* as a feature extraction technique for k-Means Clustering algorithm. The results obtained by the combination of *Fuzzy Association Rule* and *k-Means* improve the performance of text clustering by 22,04%, while combination of PCA and *k-Means* just improve the performance of text clustering by 18,05%.

Keywords: *Dimension Reduction, Principal Component Analysis, Fuzzy Association Rule, k-Means Clustering.*

PERBANDINGAN TEKNIK REDUKSI DIMENSI ANTARA ALGORITMA
PRINCIPAL COMPONENT ANALYSIS DENGAN *FUZZY ASSOCIATION RULE*
TERHADAP HASIL PENGLASTERAN

ABSTRAK

Pengklasteran merupakan proses pengelompokan data menjadi beberapa kelompok, dimana setiap anggota kelompok memiliki kesamaan dan perbedaan yang tinggi terhadap anggota kelompok lainnya. Dalam pengklasteran, algoritma konvensional bekerja baik dalam menangani data berdimensi rendah, oleh karena itu untuk meningkatkan kualitas hasil pengelompokan data teks, dibutuhkan teknik reduksi dimensi. Teknik reduksi dimensi tergolong menjadi 2 jenis, seleksi fitur dan ekstraksi fitur. Penelitian ini akan membandingkan penerapan algoritma *Principal Component Analysis* (PCA) dan *Fuzzy Association Rule* sebagai teknik reduksi ekstraksi fitur terhadap algoritma pengklasteran *k-Means*. Hasil yang didapatkan kombinasi *Fuzzy Association Rule* dan *k-Means* meningkatkan kinerja pengelompokan dokumen sebesar 22,04% dibandingkan kombinasi PCA dan *k-Means* hanya meningkatkan sebesar 18,05%.

Kata Kunci: Reduksi Dimensi, *Principal Component Analysis*, *Fuzzy Association Rule*, Pengklasteran *k-Means*.

BAB I

PENDAHULUAN

1.1 Pendahuluan

Pada bab ini akan membahas latar belakang masalah, rumusan masalah, tujuan penelitian dan manfaat penelitian serta batasan masalah. Bab ini akan memberikan penjelasan umum mengenai keseluruhan penelitian.

Pendahuluan dimulai dengan penjelasan mengenai salah satu metode dalam *data mining* yaitu pengklasteran, serta penelitian yang berkaitan dengan perbandingan ekstraksi fitur menggunakan algoritma *Principal Component Analysis* (PCA) dan algoritma *Fuzzy Association Rule* terhadap pengklasteran data berdimensi tinggi.

1.2 Latar Belakang Masalah

Dalam *data mining* terdapat banyak metode yang digunakan untuk mengolah data, salah satunya pengklasteran. Pengklasteran merupakan metode yang termasuk *unsupervised learning*, yaitu proses mengelompokkan data menjadi beberapa kelompok, dimana setiap anggota kelompok memiliki kesamaan dan perbedaan yang tinggi terhadap anggota kelompok lainnya (Han *et al*, 2012).

Dari berbagai macam algoritma dalam pengklasteran, algoritma *k-Means* merupakan algoritma yang paling terkenal, dikarenakan algoritma ini merupakan salah satu algoritma yang sederhana, sehingga lebih mudah dipahami. Selain itu, algoritma *k-Means* sangat efisien sehingga memiliki waktu pemrosesan yang lebih

cepat (Zade, Bamnote, & Agrawal, 2017; P, Sunil, & Balachandran, 2013)

Namun, pada pengelompokan data berdimensi tinggi, algoritma konvensional *k-Means* rentan terhadap data yang memiliki atribut besar sehingga menyebabkan “*curse of dimensionality*”. Oleh karena itu, mengurangi dimensi juga berarti mengurangi kompleksitas dari data tersebut. Hal yang diperlukan untuk mengatasi ini adalah dengan cara mereduksi dimensi data tersebut (Han *et al*, 2012 & Alpayidin, 2014).

Banyak keuntungan yang didapat pada reduksi dimensi, salah satunya adalah banyak algoritma *data mining* yang bekerja dengan baik jika dimensi lebih rendah. Hal ini dikarenakan, reduksi dimensi dapat menghilangkan “*curse of dimensionality*”. Alasan lain, reduksi dimensi dapat menjadikan model lebih dapat dipahami karena dibangun menggunakan fitur yang lebih sedikit. Selain itu, reduksi dimensi juga dapat mempermudah dalam visualisasi data. Bahkan jika reduksi dimensi tidak mengurangi data menjadi dua atau tiga dimensi, data sering divisualisasikan dengan mengamati pasangan dua atau tiga atribut tetapi jumlah atributnya telah banyak berkurang. Keuntungan terakhir adalah jumlah waktu dan memori yang dibutuhkan oleh algoritma *data mining* berkurang dengan adanya reduksi dimensi (Prasetyo, 2014). Teknik reduksi dimensi dalam pengklasteran itu sendiri tergolong menjadi dua yaitu seleksi atribut dan ekstraksi atribut. Seleksi atribut bertujuan untuk menemukan atribut terbaik pada dimensi dengan mendapatkan informasi dan membuang atribut yang tidak signifikan pada dimensi lain. Ekstraksi atribut bertujuan untuk mencari satu set atribut yang baru didimensi

dengan cara mengubah dimensi tinggi ke dimensi yang lebih rendah (Hasanah *et al*, 2018).

Penelitian oleh (Hasanah *et al*, 2018) mengatakan bahwa teknik reduksi dimensi dengan seleksi fitur menggunakan algoritma *Singular Value Decomposition (SVD)* lebih baik dalam hasil pengklasteran dan memiliki waktu pemrosesan lebih cepat dibandingkan teknik reduksi dimensi dengan ekstraksi fitur menggunakan algoritma PCA. Walaupun telah diuji dan mendapatkan hasil yang lebih lama dibandingkan seleksi fitur, (Han *et al*, 2012) mengatakan ekstraksi fitur masih dibutuhkan untuk pengklasteran karena tidak akan mengurangi nilai informasi dalam mereduksi data.

Selanjutnya, penelitian oleh (Rozi, Faticah and Purwitasari, 2015) mengatakan salah satu metode yang dapat mengelompokkan dokumen dengan tepat adalah menggunakan *fuzzy association rule*. Tahap ekstraksi kata kunci serta tipe *fuzzy* yang digunakan, berpengaruh terhadap kualitas pengelompokan dokumen tersebut. Penggunaan hipernim dalam ekstraksi kata kunci untuk mendapatkan suatu klaster label dapat memperluas makna dari klaster label, sehingga dapat diperoleh suatu *meaningful* klaster label, selain itu ambiguitas dan ketidakpastian yang terjadi di dalam aturan *Fuzzy Logic Systems (FLS)* tipe-1 dapat diatasi dengan *fuzzy set* tipe-2.

Berdasarkan hal tersebut, maka dalam penelitian ini akan melihat adakah algoritma yang lebih baik dari algoritma PCA dan nantinya algoritma PCA akan dibandingkan dengan algoritma *Fuzzy Association Rule*.

1.3 Rumusan Masalah

Rumusan masalah pada penelitian ini adalah bagaimana perbedaan hasil reduksi dimensi oleh algoritma PCA dibandingkan dengan algoritma *Fuzzy Association Rule* pada pengklasteran data berdimensi tinggi ? Untuk menjawab rumusan masalah tersebut, diuraikan beberapa pertanyaan penelitian sebagai berikut

1. Bagaimana metode ekstraksi fitur algoritma PCA dan *Fuzzy Association Rule* mereduksi data berdimensi tinggi ?
2. Bagaimana algoritma *k-Means* mengklaster suatu data ?
3. Apa pengaruh besaran dimensi data terhadap hasil pengklasteran ?
4. Bagaimana mengevaluasi hasil ekstraksi fitur menggunakan algoritma PCA dan *Fuzzy Association Rule* ?

1.4 Tujuan Penelitian

Tujuan dilakukannya penelitian adalah sebagai berikut :

1. Mereduksi dimensi data yang berdimensi tinggi dengan metode ekstraksi fitur menggunakan algoritma PCA dan *Fuzzy Association Rule*.
2. Mengklaster data dokumen teks dengan algoritma *k-Means*.
3. Mengevaluasi ekstraksi fitur algoritma PCA dan *Fuzzy Association Rule*.

1.5 Manfaat Penelitian

Manfaat yang dapat diperoleh dari penelitian ini adalah sebagai berikut :

1. Dapat mengklaster dokumen teks dengan baik.
2. Dapat mengoptimasi cara kerja algoritma *k-Means*.
3. Dapat mengetahui alternatif-alternatif selain algoritma PCA untuk ekstraksi fitur.

1.6 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut :

1. Metode pengklasteran yang digunakan hanya pengklasteran *k-Means*.
2. Pada penelitian ini data yang digunakan adalah dokumen teks berbahasa Indonesia yang bersumber dari situs *garuda.ristekdikti.go.id*.
3. Evaluasi kualitas pengklasteran dilakukan dengan pengukuran *Davies Bouldin Index (DBI)*.

1.7 Kesimpulan

Penelitian ini berfokus pada bagaimana perbandingan antara algoritma PCA dan *Fuzzy Association Rule* dalam ekstraksi atribut data berdimensi tinggi terhadap hasil pengklasteran. Seperti yang dijelaskan pada BAB I, dalam rumusan masalah dikatakan bahwa teknik reduksi akan menghasilkan pengelompokkan dengan lebih baik. Data berdimensi tinggi akan direduksi menggunakan algoritma

PCA dan *Fuzzy Association Rule*. Penelitian ini akan melihat manakah algoritma yang lebih baik dalam mereduksi dimensi. Setelah direduksi, selanjutnya data akan dilakukan pengklasteran menggunakan metode *k-Means*. Kemudian setelah mendapat hasil pengklasteran, maka akan dievaluasi menggunakan evaluasi internal *Davies Bouldin Index* (DBI). Dalam membandingkan hasil pengklasteran digunakan kondisi di mana semakin kecil nilai DBI yang dihasilkan maka semakin baik pula hasil pengklasteran yang dihasilkan.

DAFTAR PUSTAKA

- Alpaydin, E., Introduction to machine learning. 2014: MIT press.
- Chen, C., F. Tseng dan T. Liang (2010). "Mining Fuzzy Frequent Itemset For Hierarchical Document Clustering". *Journal of Information Processing and Management*, vol. 46, 2:193-211.
- Cohen, L., Manion, L., & Morrison, K. 2007. Research Methods in Education (6th ed.). London, New York: Routledge Falmer
- Han, J., M. Kamber and J. Pei (2012). "Data Mining Concept and Techniques (3rd Edition)".
- Hasanah, S. I. R., M. I. Jambak, D. M. Saputra. (2018). 'Comparison of Dimensional Reduction Using Singular Value Decompositions and Principal Component Analysis for Clustering Result for Indonesian Language Text Document'. Faculty of Computer Science, Sriwijaya University, Palembang, Indonesia.
- Hidayati, Jambak dan Saputra (2017) 'Pengaruh Singular Value Decomposition Terhadap Metode – Metode Clustering', pp. 95–104.
- Kruchten, P. (2000). *The Rational Unified Process An Introduction, Second Edition*. United States of America: Addison Wesley Longman, Inc.
- Nazief, B. A. and M. Adriani (1996). "Confix Stripping: Approach to Stemming Algorithm for Bahasa Indonesia." Internal publication, Faculty of Computer Science, University of Indonesia, Depok, Jakarta 41.
- Ma'arif, Abdul Aziz. 2015. Penerapan Algoritma Tf-Idf Untuk Pencarian Karya Ilmiah. Jurnal. Jurusan Teknik Informatika. Fakultas Ilmu Komputer. Universitas Dian Nuswantoro. Semarang.
- Melita, R. et al. (2018) Penerapan Metode Term Frequency Inverse Document Frequency (TF-IDF) Dan Cosine Similarity Pada Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits Berbasis Web', *Jurna Teknik Informatika*, Vol 11 No. 2.
- Merliana, N. P. E., & Santoso, A. J. (2015). Analisa Penentuan Jumlah Cluster Terbaik pada Metode K-Means Clustering.
- P, B., Sunil, K. P. V and Balachandran, K. P. (2013) 'K-Means Clustering - Review of Various Methods for Initial Selection of Centroids', 4(8), p. 4.
- Prasetyo, E. (2014, September). Reduksi Dimensi Set Data dengan DRC pada

Metode Klasifikasi SVM dengan Upaya Penambahan Komponen Ketiga. In *Seminar Nasional Teknologi dan Informatika 2014*. Muria Kudus University.

Rahayu, G., & Mustakim, M. (2017, May). *Principal Component Analysis* untuk Dimensi Reduksi Data *Clustering* Sebagai Pemetaan Persentase Sertifikasi Guru di Indonesia. In *Seminar Nasional Teknologi Informasi Komunikasi dan Industri* (pp. 201-208).

Ristevski, B., S. Loshkovska, S. Dzeroski and I. Slavkov (2008). "A Comparison of Validation Indices for Evaluation of Clustering Results of DNA Microarray Data." *The 2nd International Conference on Bioinformatics and Biomedical Engineering (ICBBE)*: 587-591.

Roy, A., & Chatterjee, R. (2013). *A survey on fuzzy association rule mining methodologies*. *IOSR J. Comput. Eng.(IOSR-JCE)*, e-ISSN, 2278-0661.

Rozi, F., Chastine Fatichah, and Diana Purwitasari. (2015). "Ekstraksi Kata Kunci Berdasarkan Hipernim dengan Inisialisasi *Cluster* Menggunakan *Fuzzy Association Rule Mining* pada Pengelompokan Dokumen ", *Jurnal Teknologi Informasi (JUTI)*, vol. 13, no. 2, pp. 190-197.

Santosa, Budi dan Ardian Umam. 2018. *Data Mining dan Big Data Analytics*. Yogyakarta : Penebar Media Pustaka.

Younus, Z. S., Mohamad, D., Saba, T., Alkawaz, M. H., Rehman, A., Al-Rodhaan, M., & Al-Dhelaan, A. (2014). Content-based image retrieval using PSO and *K-Meansclustering* algorithm. *Arabian Journal of Geosciences*, (Salamah 2010). <https://doi.org/10.1007/s12517-014-1584-7>

Zade, J., Bamnote, G. R. and Agrawal, P. P. K. (2017) '*Text Document Clustering Using K-means Algorithm with Its Analysis and Implementation*', (2).