

***CLUSTERING ARTIKEL ILMIAH BERBAHASA INDONESIA
MENGUNAKAN ALGORITMA PARTITIONING AROUND
MEDOIDS***

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



Oleh:

**PRISTI KARTIKA SURYANI
NIM: 09021181924017**

**Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2024**

LEMBAR PENGESAHAN SKRIPSI

CLUSTERING ARTIKEL ILMIAH BERBAHASA INDONESIA MENGUNAKAN ALGORITMA PARTITIONING AROUND MEDOIDS

Oleh:

PRISTI KARTIKA SURYANI
NIM: 09021181924017

Indralaya, 24 April 2024
Pembimbing II,

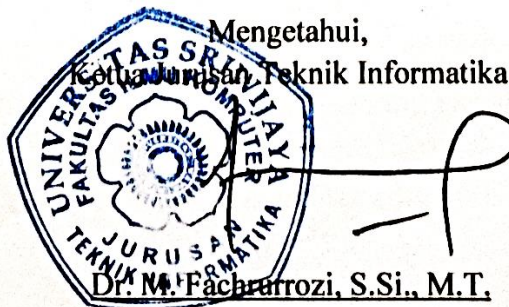
Pembimbing I,



Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003



M. Naufal Rachmatullah, M.T.
NIP. 199212012022031008



NIP. 198005222008121002

TANDA LULUS UJIAN SIDANG KOMPREHENSIF

Pada hari Jumat tanggal 5 April 2024 telah dilaksanakan Ujian Komprehensif Skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Pristi Kartika Suryani
NIM : 09021181924017
Judul : *Clustering* Artikel Ilmiah Berbahasa Indonesia Menggunakan Algoritma *Partitioning Around Medoids*

dan dinyatakan LULUS.

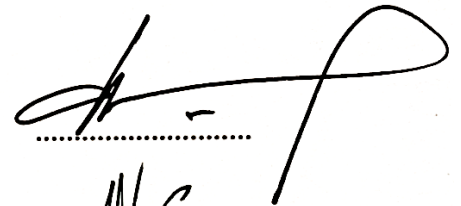
1. Ketua

Mastura Diana Marieska, M.T.
NIP. 198603212018032001



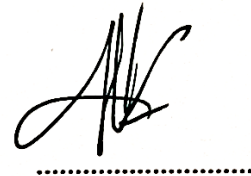
2. Penguji

Julian Supardi, M.T., Ph.D.
NIP. 197207102010121001



3. Pembimbing I

Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

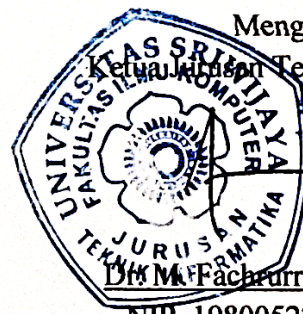


4. Pembimbing II

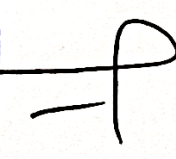
Muhammad Naufal Rachmatullah, M.T.
NIP. 199212012022031008



Mengetahui,
Ketua Jurusan Teknik Informatika



Drs. M. Fachrudzi, S.Si., M.T.
NIP. 198005222008121002



HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Pristi Kartika Suryani

NIM : 09021181924017

Program Studi : Teknik Informatika

Judul Skripsi : *Clustering* Artikel Ilmiah Berbahasa Indonesia Menggunakan
Algoritma *Partitioning Around Medoids*

Hasil pengecekan *Software iThenticate/Turnitin* : **17%**

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

Demikian pernyataan ini saya buat dengan sebenarnya tanpa paksaan dari pihak manapun.



Palembang, 24 April 2024



Pristi Kartika Suryani
NIM. 09021181924017

MOTTO DAN PERSEMBAHAN

"I am a slow walker, but I never walk back."

(Abraham Lincoln)

"Dibalik seseorang yang memiliki kekurangan, pasti terdapat kelebihan yang Tuhan anugerahkan."

(Pristi Kartika)

Saya dedikasikan karya tulis ini kepada:

- Ibu, Bapak, dan separuh hatiku, separuh jiwaku... Kirana, Adikku tersayang...
- Bu Alvi dan Pak Naufal, sebagai Dosen Pembimbing Terbaik
- Jurusan Teknik Informatika
- Fakultas Ilmu Komputer
- Universitas Sriwijaya

ABSTRACT

A scientific article is a type of written work by a researcher after successfully conducting research in a particular field, which is arranged systematically based on writing rules. Each scientific article contains important information from various scientific fields, one of which is computer science. In this research, an automatic clustering software for scientific articles in Indonesian was built, where the data used was the abstract of each computer science scientific article. Clustering is a method for grouping objects into several clusters so that each cluster contains objects that are similar based on certain categories. In this research, the clustering process was carried out to find out words that often and rarely appear in computer science scientific articles. The methods used in software development begin with the text preprocessing process, word weighting with TF-IDF, searching for the best k value with Silhouette Score, and the clustering process with the Partitioning Around Medoids algorithm, better known as k -Medoids. The experimental results of the software that has been built were the formation of 10 optimal clusters in the first experiment using the best k value ($k=10$ with a Silhouette Score of 0.7123), and the formation of 1 cluster in the second experiment using $k=2$ (with a Silhouette Score of 0). This software also has good performance, as seen in every experimental process carried out.

Keywords: clustering, scientific article, k-medoids

Advisor I,



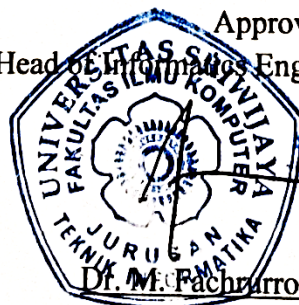
Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

Advisor II,



M. Naufal Rachmatullah, M.T.
NIP. 199212012022031008

Approved,
Head of Informatics Engineering Department



Df. M. Fachrudrozi, S.Si., M.T.
NIP. 198005222008121002

ABSTRAK

Artikel ilmiah adalah tulisan yang dibuat oleh seorang peneliti setelah berhasil melakukan penelitian di bidang tertentu, yang disusun secara sistematis sesuai dengan kaidah penulisan. Setiap artikel ilmiah mengandung informasi yang dibutuhkan dari berbagai bidang keilmuan, salah satunya bidang ilmu komputer. Dalam penelitian ini, dibangun sebuah perangkat lunak *clustering* artikel ilmiah berbahasa Indonesia secara otomatis, di mana data yang digunakan merupakan abstrak dari setiap artikel ilmiah ilmu komputer. *Clustering* merupakan metode untuk mengelompokkan objek ke dalam beberapa *cluster*, sehingga dalam setiap *cluster* akan berisi objek yang memiliki kemiripan berdasarkan kategori tertentu. Pada penelitian ini, proses *clustering* tersebut dilakukan untuk mengetahui kata-kata yang sering dan yang jarang sekali muncul dalam sebuah dokumen artikel ilmiah ilmu komputer. Adapun metode-metode yang digunakan dalam pembangunan perangkat lunak diawali dengan proses *text preprocessing*, pembobotan kata menggunakan TF-IDF, pencarian nilai k terbaik menggunakan *Silhouette Score*, dan proses *clustering* menggunakan algoritma *Partitioning Around Medoids* atau lebih dikenal dengan sebutan *k-Medoids*. Hasil percobaan dari perangkat lunak yang telah dibangun adalah terbentuknya 10 *cluster* optimal pada percobaan pertama menggunakan nilai k terbaik ($k=10$ dengan *Silhouette Score* sebesar 0.7123), dan terbentuknya 1 *cluster* pada percobaan kedua menggunakan $k=2$ (dengan *Silhouette Score* sebesar 0). Perangkat lunak tersebut juga memiliki kinerja yang baik, dilihat dari setiap proses percobaan yang dilakukan.

Kata Kunci: *clustering*, artikel ilmiah, *k-medoids*

Pembimbing I,



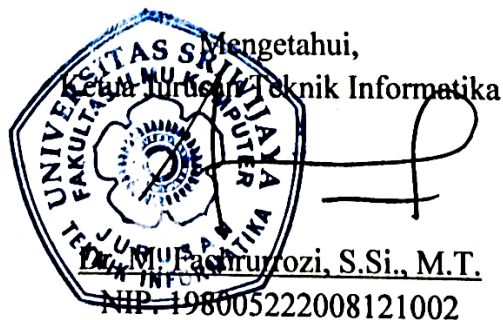
Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

Pembimbing II,



M. Naufal Rachmatullah, M.T.
NIP. 199212012022031008

Mengetahui,
Kepala Jurusan Teknik Informatika



M. Fachrudzi, S.Si., M.T.
NIP. 198005222008121002

KATA PENGANTAR

Puji dan syukur Penulis panjatkan ke hadirat Tuhan Yang Maha Esa, karena berkat rahmat, berkat, dan karunia-Nya, Penulis dapat menyelesaikan Laporan Skripsi ini dengan baik. Skripsi ini disusun guna memenuhi syarat untuk menyelesaikan Pendidikan Program Strata-1 pada Jurusan Teknik Informatika, Fakultas Ilmu Komputer di Universitas Sriwijaya.

Dalam menyelesaikan Laporan Skripsi ini, banyak pihak yang telah memberikan bantuan, dukungan, nasihat, dan bimbingan, baik secara langsung maupun tidak langsung. Oleh karena itu, dalam kesempatan ini Penulis ingin mengucapkan ribuan terima kasih banyak kepada:

1. Orang Tua tersayang, orang tua terhebat dan terbaik di dunia, Pak Black Hunter dan Ibu Pris, Adik tersayang, adik terbijaksana dan terbaik di dunia, Surya Kirana, keluarga luar biasa yang selalu ada dan selalu menjadi *support system* dalam segala situasi dan kondisi hidup Penulis, yang selalu mendo'akan dan memberikan yang terbaik untuk hidup Penulis, khususnya dalam penyusunan Skripsi ini.
2. Bapak Prof. Dr. Erwin, S.Si., M.Si. selaku Dekan Fasilkom Universitas Sriwijaya.
3. Bapak Julian Supardi, S.Pd., M.T., Ph.D. selaku Wakil Dekan I Bidang Akademik Fasilkom Universitas Sriwijaya sekaligus sebagai Dosen Penguji yang telah memberikan saran dan masukan terbaik untuk Penulis.
4. Bapak Dr. Muhammad Fachrurrozi, S.Si., M.T. selaku Ketua Jurusan Teknik Informatika Fasilkom Universitas Sriwijaya.
5. Ibu Alvi Syahrini Utami, S.Si., M.Kom. selaku Dosen Pembimbing Akademik sekaligus Dosen Pembimbing I Skripsi, yang selalu mengarahkan dan membimbing Penulis selama penyusunan Skripsi ini.
6. Bapak Muhammad Naufal Rachmatullah, S.Kom., M.T. selaku Dosen Pembimbing II Skripsi, yang selalu mengarahkan dan membimbing Penulis selama penyusunan Skripsi ini.
7. Ibu Mastura Diana Marieska, S.T., M.T. selaku Dosen Teknik Informatika sekaligus sebagai Ketua Penguji yang senantiasa mengarahkan Penulis.
8. Bapak Rifkie Primartha, S.T., M.T. selaku Dosen Teknik Informatika yang selalu memberikan wejangan terbaik selama Penulis menjadi mahasiswa di Fasilkom, dan senantiasa membantu Penulis selama penyusunan Skripsi ini.
9. Seluruh Dosen Fasilkom Universitas Sriwijaya, khususnya Dosen Jurusan Teknik Informatika.
10. Kak Ricy Firnando, S.Kom., M.Kom. selaku Admin Jurusan Teknik Informatika yang terdahulu, dan Mba Septy Widyastuti, S.E. selaku Admin Jurusan Teknik Informatika yang sekarang, yang sudah banyak membantu Penulis dalam urusan administrasi dan akademik selama masa perkuliahan.
11. Helmi, teman sebangku pertama di hari pertama Penulis masuk ke kelas TIREG C, sobat terbaik yang dari dulu sampai sekarang selalu memberikan semangat, bantuan, dan dukungan terbaiknya untuk Penulis. Do'a terbaik selalu menyertai kamu, Mi.

12. Khoir, sobat seperjuangan Penulis yang selalu ada. Sobat terbaik yang selalu mendukung, menyemangati, dan tetap kebersamai Penulis selama ini; disaat semuanya terasa amat sangat kalut dan tak menentu. Do'a terbaik selalu menyertai kamu, Khoir.
 13. Rifqi, sobat lawak yang tiba-tiba. Sobat Maxwell x Karen; sobat bincang Penulis yang turut serta memberikan semangat, dukungan, dan bantuan yang tiba-tiba; tak terduga. Do'a terbaik selalu menyertai kamu, Qi.
 14. Aditz Friz, sobat ribut Penulis yang selalu ada dan tiada. Sobat bincang sana bincang sini yang turut serta mendukung, menyemangati, dan tetap kebersamai Penulis selama ini walaupun banyak ributnya. Do'a terbaik selalu menyertai kamu, Ditz.
 15. Luthfi, sobat terbaik Penulis yang selalu memberikan saran, masukan, semangat, dan dukungan yang luar biasa selama Penulis menjalani suka duka kehidupan perkuliahan. Do'a terbaik selalu menyertai kamu, Fi.
- Penulis menyadari bahwa Skripsi ini masih jauh dari kata sempurna. Oleh karena itu, kritik dan saran yang bersifat membangun sangat diperlukan untuk menyempurnakan Skripsi ini supaya lebih bermanfaat bagi kita semua.

Palembang, 24 April 2024



Pristi Kartika Suryani

DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PENGESAHAN SKRIPSI	ii
TANDA LULUS UJIAN SIDANG KOMPREHENSIF.....	iii
HALAMAN PERNYATAAN.....	iv
MOTTO DAN PERSEMBAHAN.....	v
ABSTRACT	vi
ABSTRAK	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR GAMBAR.....	xiv
DAFTAR TABEL	xvi
BAB I PENDAHULUAN.....	I-1
1.1. Pendahuluan	I-1
1.2. Latar Belakang Masalah.....	I-1
1.3. Rumusan Masalah	I-3
1.4. Tujuan Penelitian	I-4
1.5. Manfaat Penelitian	I-4
1.6. Batasan Masalah.....	I-4
1.7. Sistematika Penulisan	I-5
1.8. Kesimpulan	I-6
BAB II KAJIAN LITERATUR	II-1
2.1. Pendahuluan	II-1
2.2. Landasan Teori.....	II-1
2.2.1 <i>Clustering</i>	II-1
2.2.2 <i>Text Pre-processing</i>	II-2
a. <i>Case Folding</i>	II-2
b. <i>Tokenizing</i>	II-3
c. <i>Stopword Removal (Filtering)</i>	II-3

d. <i>Stemming</i>	II-4
2.2.3 Pembobotan TF-IDF	II-6
2.2.4 <i>Partitioning Around Medoids (k-Medoids)</i>	II-7
2.2.5 <i>Silhouette Score</i>	II-9
2.2.6 Metode <i>RUP (Rational Unified Process)</i>	II-10
a. Fase Insepsi (<i>Inception Phase</i>).....	II-10
b. Fase Elaborasi (<i>Elaboration Phase</i>).....	II-10
c. Fase Konstruksi (<i>Construction Phase</i>).....	II-11
d. Fase Transisi (<i>Transition Phase</i>)	II-11
2.3. Penelitian Lain yang Relevan.....	II-11
2.4. Kesimpulan	II-13
BAB III METODOLOGI PENELITIAN	III-1
3.1. Pendahuluan	III-1
3.2. Unit Penelitian.....	III-1
3.3. Metode Pengumpulan Data	III-1
3.3.1 Jenis Data	III-1
3.3.2 Teknik Pengumpulan Data	III-1
3.4. Langkah-Langkah Penelitian	III-2
3.4.1 Melakukan Studi Literatur Algoritma <i>PAM</i>	III-3
3.4.2 Melakukan Pengumpulan Data	III-3
3.4.3 Melakukan Pengembangan Perangkat Lunak	III-4
3.4.4 Melakukan Pengujian Perangkat Lunak & Menganalisis Hasil	III-5
3.5. Metode Pengembangan Perangkat Lunak.....	III-5
a. Fase Insepsi (<i>Inception Phase</i>).....	III-5
b. Fase Elaborasi (<i>Elaboration Phase</i>).....	III-5
c. Fase Konstruksi (<i>Construction Phase</i>).....	III-6
d. Fase Transisi (<i>Transition Phase</i>)	III-6
BAB IV PEMBANGUNAN PERANGKAT LUNAK	IV-1
4.1. Pendahuluan	IV-1
4.2. Fase Insepsi (<i>Inception Phase</i>)	IV-1
4.2.1 Pemodelan Bisnis	IV-1

4.2.2	Analisis Kebutuhan (<i>Requirement Analysis</i>)	IV-2
	a. Kebutuhan Fungsional (<i>Functional Requirement</i>)	IV-2
	b. Kebutuhan Non-Fungsional (<i>Non-Functional Requirement</i>)....	IV-3
4.2.3	Analisis Data (<i>Data Analysis</i>)	IV-3
4.2.4	<i>Use Case Diagram</i>	IV-8
	4.2.4.1 Definisi Aktor	IV-9
	4.2.4.2 Definisi <i>Use Case</i>	IV-10
	4.2.4.3 Skenario <i>Use Case</i>	IV-11
4.3.	Fase Elaborasi (<i>Elaboration Phase</i>)	IV-15
	4.3.1 Pemodelan Bisnis	IV-15
	4.3.2 Perancangan Data	IV-16
	4.3.3 Perancangan Antarmuka (<i>Interface</i>)	IV-16
	4.3.4 <i>Activity Diagram</i>	IV-18
	4.3.5 <i>Sequence Diagram</i>	IV-21
4.4.	Fase Konstruksi (<i>Construction Phase</i>)	IV-22
	4.4.1 Kebutuhan Sistem	IV-22
	4.4.2 <i>Class Diagram</i>	IV-23
	4.4.3 Implementasi	IV-24
	4.4.3.1 Lingkungan Implementasi.....	IV-24
	4.4.3.2 Implementasi Kelas	IV-25
	4.4.3.3 Implementasi <i>Interface</i>	IV-26
4.5.	Fase Transisi (<i>Transition Phase</i>)	IV-27
	4.5.1 Pemodelan Bisnis	IV-27
	4.5.2 Rencana Pengujian	IV-27
	4.5.2.1 Rencana Pengujian Pada Tahap <i>Preprocessing</i>	IV-28
	4.5.2.2 Rencana Pengujian Pada Tahap <i>Silhouette Score</i>	IV-28
	4.5.2.3 Rencana Pengujian Pada Tahap <i>Clustering</i>	IV-28
	4.5.3 Pengujian	IV-29
	4.5.3.1 Pengujian Pada Tahap <i>Preprocessing</i>	IV-29
	4.5.3.2 Pengujian Pada Tahap <i>Silhouette Score</i>	IV-30
	4.5.3.3 Pengujian Pada Tahap <i>Clustering</i>	IV-31

4.6.	Kesimpulan	IV-32
BAB V HASIL DAN ANALISIS PENELITIAN.....		V-1
5.1.	Pendahuluan	V-1
5.2.	Data Hasil Penelitian.....	V-1
5.2.1	Konfigurasi Percobaan	V-1
5.2.2	Data Hasil Konfigurasi.....	V-8
5.3.	Hasil dan Analisis Penelitian	V-8
5.4.	Kesimpulan	V-14
BAB VI KESIMPULAN DAN SARAN.....		VI-1
6.1.	Kesimpulan	VI-1
6.2.	Saran.....	VI-2
DAFTAR PUSTAKA		xvii

DAFTAR GAMBAR

Gambar II-1 Contoh Hasil <i>Case Folding</i>	II-2
Gambar II-2 Contoh Hasil <i>Tokenizing</i>	II-3
Gambar II-3 Contoh Hasil <i>Stopword Removal (Filtering)</i>	II-4
Gambar II-4 Contoh Hasil <i>Stemming</i>	II-4
Gambar III-1 Diagram Alur Penelitian	III-3
Gambar III-2 Arsitektur Perangkat Lunak	III-4
Gambar IV-1 Hasil <i>Tokenizing</i>	IV-5
Gambar IV-2 Hasil <i>Stopword Removal</i>	IV-5
Gambar IV-3 Hasil <i>Stemming</i>	IV-6
Gambar IV-4 Tahap Pembobotan Kata (TF-IDF)	IV-7
Gambar IV-5 <i>Use Case Diagram</i>	IV-8
Gambar IV-6 Rancangan Antarmuka <i>Preprocessing</i>	IV-16
Gambar IV-7 Rancangan Antarmuka <i>Silhouette Score</i>	IV-17
Gambar IV-8 Rancangan Antarmuka <i>Clustering</i>	IV-17
Gambar IV-9 <i>Activity Diagram</i> 001	IV-18
Gambar IV-10 <i>Activity Diagram</i> 002	IV-19
Gambar IV-11 <i>Activity Diagram</i> 003	IV-20
Gambar IV-12 <i>Sequence Diagram</i> 001	IV-21
Gambar IV-13 <i>Sequence Diagram</i> 002	IV-21
Gambar IV-14 <i>Sequence Diagram</i> 003	IV-22
Gambar IV-15 <i>Class Diagram</i>	IV-23
Gambar V-1 Isi Data 3	V-2
Gambar V-2 Hasil <i>Case Folding</i> Data 3	V-2
Gambar V-3 Hasil <i>Tokenizing</i> Data 3	V-3
Gambar V-4 Hasil <i>Filtering</i> Data 3	V-4
Gambar V-5 Hasil <i>Stemming</i> Data 3	V-5
Gambar V-6 Data 3 Pasca Tahap <i>Text Preprocessing</i>	V-5
Gambar V-7 Contoh Pertama Hasil Pembobotan Kata	V-6
Gambar V-8 Contoh Kedua Hasil Pembobotan Kata	V-6
Gambar V-9 Hasil <i>Silhouette Score</i>	V-7
Gambar V-10 Hasil <i>Clustering</i>	V-8
Gambar V-11 Hasil <i>Clustering</i> 2	V-9
Gambar V-12 Hasil <i>Clustering</i> 3	V-10
Gambar V-13 Hasil <i>Clustering</i> 4	V-10
Gambar V-14 Hasil <i>Clustering</i> 5	V-11
Gambar V-15 Hasil <i>Clustering</i> 6	V-11

Gambar V-16 Visualisasi Hasil 10 <i>Cluster</i>	V-12
Gambar V-17 Keterangan Nilai Visualisasi Hasil.....	V-12
Gambar V-18 Visualisasi Hasil 1 <i>Cluster</i>	V-14

DAFTAR TABEL

Tabel IV-1 Kebutuhan Fungsional Perangkat Lunak.....	IV-2
Tabel IV-2 Kebutuhan Non-Fungsional Perangkat Lunak	IV-3
Tabel IV-3 Definisi Aktor	IV-9
Tabel IV-4 Definisi <i>Use Case</i>	IV-10
Tabel IV-5 Skenario <i>Use Case</i> 001	IV-11
Tabel IV-6 Skenario <i>Use Case</i> 002.....	IV-13
Tabel IV-7 Skenario <i>Use Case</i> 003.....	IV-14
Tabel IV-8 Implementasi Kelas	IV-25
Tabel IV-9 Rencana Pengujian <i>Use Case</i> 001	IV-28
Tabel IV-10 Rencana Pengujian <i>Use Case</i> 002	IV-28
Tabel IV-11 Rencana Pengujian <i>Use Case</i> 003	IV-28
Tabel IV-12 Hasil Pengujian <i>Use Case</i> 001	IV-29
Tabel IV-13 Hasil Pengujian <i>Use Case</i> 002	IV-30
Tabel IV-14 Hasil Pengujian <i>Use Case</i> 003	IV-31

BAB I

PENDAHULUAN

1.1. Pendahuluan

Pada bab ini diuraikan mengenai latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, sistematika penulisan, dan kesimpulan dari bab Pendahuluan.

1.2. Latar Belakang Masalah

Artikel ilmiah adalah tulisan yang dibuat oleh seorang peneliti setelah berhasil melakukan penelitian di bidang tertentu, yang disusun secara sistematis sesuai dengan kaidah penulisan. Menurut Komara (2017), artikel ilmiah merupakan sebuah tulisan faktual atau non-fiksi tentang masalah tertentu yang dimuat dalam jurnal, majalah, dan buletin guna menyampaikan gagasan dan fakta, meyakinkan, mendidik, dan menawarkan solusi dari suatu masalah. Di dunia perkuliahan, artikel ilmiah tersebut biasa dikenal dengan istilah *paper*. Artikel ilmiah yang sudah dipublikasikan biasanya digunakan sebagai rujukan oleh setiap orang sebelum melakukan penelitian yang baru, karena pada setiap artikel ilmiah biasanya mengandung informasi yang dibutuhkan dari berbagai bidang keilmuan. Pada penelitian ini, data yang akan digunakan adalah artikel-artikel ilmiah berbahasa Indonesia di bidang ilmu komputer. Artikel ilmiah tersebut akan digunakan dalam proses pengelompokan tak berlabel, atau disebut dengan proses *clustering*, namun dirancang sedemikian rupa agar dapat dilakukan secara otomatis, dengan tujuan

untuk mengetahui kata-kata yang sering dan yang jarang sekali muncul dalam sebuah artikel ilmiah ilmu komputer. Oleh karena itu, untuk mewujudkan hal tersebut, pada penelitian ini akan dikembangkan sebuah perangkat lunak *clustering* artikel ilmiah berbahasa Indonesia secara otomatis menggunakan algoritma atau metode yang dapat diimplementasikan pada proses *clustering*.

Clustering merupakan suatu metode untuk mengelompokkan data ke dalam beberapa *cluster*, sehingga dalam setiap *cluster* akan berisi data yang memiliki kemiripan, dan akan berbeda dengan data yang ada dalam *cluster* lainnya. *Cluster* berbeda dengan *group*. *Group* berisi data yang memiliki kesamaan kondisi, sedangkan *cluster* berisi data yang memiliki kemiripan berdasarkan pada kedekatan suatu karakteristik dari *sample* yang diberikan. Adapun perbedaan antara *clustering* dengan *classification* adalah tidak adanya variabel target dalam melakukan pengelompokan data pada proses *clustering* (Pramesti, Furqon, dan Dewi, 2017). Salah satu algoritma *clustering* adalah *Partitioning Around Medoids* (PAM Algorithm). *Partitioning Around Medoids* merupakan salah satu algoritma *clustering* yang berfungsi untuk melakukan partisi pada *dataset*, di mana algoritma ini menggunakan objek sebagai perwakilan atau pusat dari sebuah *cluster* yang disebut dengan *medoid*. *Partitioning Around Medoids* lebih dikenal dengan sebutan *k-Medoids* (Wibawa, Miftahuddin, dan Suyono, 2021).

Sebagai salah satu algoritma *clustering*, algoritma *k-Medoids* merupakan bentuk pengembangan variasi dari algoritma *k-Means* (Maharani, Ratnawati, dan Rahayudi, 2021). Penelitian terdahulu yang menggunakan algoritma *k-Medoids* pada proses *clustering* telah dilakukan oleh Wibawa, Miftahuddin, dan Suyono

pada tahun 2021. Penelitian tersebut melakukan *clustering* untuk pembentukan *database stopwords* bahasa Jawa. Pada penelitian tersebut, setiap hasil *cluster* diuji dengan melakukan pencocokan terhadap *stopword* hasil identifikasi ahli bahasa Jawa. Hasilnya, penelitian tersebut menunjukkan bahwa *stopword* yang dihasilkan menggunakan *k-Medoids clustering* dengan nilai $k=13$ memiliki hasil pengujian sebesar 70.5%. Penelitian terkait lainnya dilakukan oleh Soni dan Patel pada tahun 2017. Pada penelitian ini, penulis melakukan perbandingan antara algoritma *k-Means* dan *k-Medoids* menggunakan *dataset* tanaman Iris yang diperoleh dari *UCI Machine Learning Repository*. Hasil yang diperoleh menunjukkan bahwa algoritma *k-Medoids* memiliki kemampuan yang lebih baik dalam skalabilitas untuk *dataset* yang lebih besar, dan juga lebih efisien daripada algoritma *k-Means* yang sensitif terhadap *outlier*. Dari beberapa penelitian terkait yang pernah dilakukan, maka pada penelitian ini akan dikembangkan sebuah perangkat lunak *clustering* artikel ilmiah berbahasa Indonesia secara otomatis menggunakan algoritma *Partitioning Around Medoids* atau *k-Medoids*.

1.3. Rumusan Masalah

Rumusan masalah dari penelitian ini adalah sebagai berikut:

1. Bagaimana *clustering* artikel ilmiah berbahasa Indonesia secara otomatis menggunakan algoritma *Partitioning Around Medoids*?
2. Bagaimana kinerja dan hasil dari perangkat lunak *clustering* artikel ilmiah berbahasa Indonesia secara otomatis yang dikembangkan menggunakan algoritma *Partitioning Around Medoids*?

1.4. Tujuan Penelitian

Tujuan penelitian ini adalah sebagai berikut:

1. Membangun perangkat lunak *clustering* artikel ilmiah berbahasa Indonesia secara otomatis menggunakan algoritma *Partitioning Around Medoids*.
2. Mengetahui kinerja dan hasil dari perangkat lunak *clustering* artikel ilmiah berbahasa Indonesia secara otomatis yang dikembangkan menggunakan algoritma *Partitioning Around Medoids*.

1.5. Manfaat Penelitian

Manfaat penelitian ini adalah sebagai berikut:

1. Terbentuknya perangkat lunak *clustering* artikel ilmiah berbahasa Indonesia secara otomatis menggunakan algoritma *Partitioning Around Medoids*.
2. Kinerja dan hasil dari perangkat lunak *clustering* artikel ilmiah berbahasa Indonesia secara otomatis menggunakan algoritma *Partitioning Around Medoids* dapat dijadikan sebagai referensi untuk penelitian di masa mendatang

1.6. Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Penelitian ini hanya dilakukan pada artikel ilmiah berbahasa Indonesia.

2. Data yang digunakan dalam penelitian ini berupa artikel ilmiah dengan format **.pdf* yang akan dikonversikan ke dalam format **.txt*, kemudian dijadikan satu menjadi sebuah *dataset* dalam format **.csv*.
3. Artikel ilmiah yang akan dikelompokkan adalah artikel ilmiah yang berkaitan dengan bidang ilmu komputer.
4. Penelitian ini hanya menggunakan bagian abstrak dari setiap artikel ilmiah.

1.7. Sistematika Penulisan

Sistematika penulisan skripsi ini adalah sebagai berikut:

BAB I. PENDAHULUAN

Pada bab ini diuraikan mengenai latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, sistematika penulisan, dan kesimpulan dari bab Pendahuluan.

BAB II. KAJIAN LITERATUR

Pada bab ini dibahas mengenai landasan-landasan teori yang digunakan sebagai dasar dalam penelitian, termasuk juga penelitian-penelitian terdahulu yang terkait, serta metode pengembangan perangkat lunak yang akan digunakan dalam penelitian.

BAB III. METODOLOGI PENELITIAN

Pada bab ini dibahas mengenai tahapan-tahapan yang akan dilaksanakan dalam penelitian. Masing-masing rencana tahapan penelitian dideskripsikan dengan

rinci, dan mengacu pada suatu kerangka kerja. Di bagian akhir bab ini berisi perancangan manajemen proyek pada pelaksanaan penelitian.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Pada bab ini dibahas mengenai analisis kebutuhan perangkat lunak, perancangan perangkat lunak, implementasi perangkat lunak, dan pengujian perangkat lunak.

BAB V. HASIL DAN ANALISIS PENELITIAN

Pada bab ini, hasil pengujian berdasarkan langkah-langkah yang telah direncanakan disajikan. Analisis diberikan sebagai basis dari kesimpulan yang diambil dalam penelitian ini.

BAB VI. KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari uraian-uraian pada bab sebelumnya, serta mengandung saran yang dapat digunakan sebagai acuan untuk melakukan pembaruan pada penelitian selanjutnya sesuai dengan perkembangan di masa mendatang.

1.8. Kesimpulan

Pada Bab I ini dapat disimpulkan bahwa dalam penelitian ini akan dikembangkan sebuah perangkat lunak *clustering* artikel ilmiah berbahasa Indonesia secara otomatis menggunakan algoritma *Partitioning Around Medoids*.

DAFTAR PUSTAKA

- Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., & Kochut, K. (2017). A brief survey of text mining: Classification, clustering and extraction techniques. *arXiv preprint arXiv:1707.02919*.
- Chen, W., Xu, Z., Zheng, X., Yu, Q., & Luo, Y. (2020). Research on sentiment classification of online travel review text. *Applied Sciences*, *10*(15), 5275.
- Chrisnanto, Y. H., & Abdillah, G. (2015). Gambaran Umum Kemampuan Akademik Mahasiswa Unjani Dengan Algoritma Partitioning Around Medoids (PAM) Clustering. In *Seminar Nasional Ilmu Pengetahuan dan Teknologi* (pp. 285-290).
- Dewi, D. A. I. C., & Pramita, D. A. K. (2019). Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali. *Matrix: Jurnal Manajemen Teknologi dan Informatika*, *9*(3), 102-109.
- Herviany, M., Delima, S. P., Nurhidayah, T., & Kasini, K. (2021). Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Daerah Rawan Tanah Longsor Pada Provinsi Jawa Barat: Comparison of K-Means and K-Medoids Algorithms for Grouping Landslide Prone Areas in West Java Province. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *1*(1), 34-40.
- Hutahaean, D. J., Wardani, N. H., & Purnomo, W. (2019). Pengembangan Sistem Informasi Penyewaan Gedung Berbasis Web dengan Metode Rational Unified Process (RUP)(Studi Kasus: Wisma Rata Medan). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN*, *2548*, 964X.
- Komara, A. (2017). Menulis Artikel dan Karya Ilmiah.
- Kruchten, P. (2004). *The rational unified process: an introduction*. Addison-Wesley Professional.
- Kurnia, H., Zahrotun, L., & Linarti, U. (2021). Pengelompokan Mahasiswa Berdasarkan Data Akademik Sebelum Kuliah dan Masa Studi Menggunakan K-Medoids. *InfoTekJar: Jurnal Nasional Informatika dan Teknologi Jaringan*, *5*(2), 265-272.
- Lapiza, R., Amalita, N., & Fitria, D. (2023). Grouping The Districts in Sumatera Region Based on Economic Development Indicators Using K-Medoids and CLARA Methods. *UNP Journal of Statistics and Data Science*, *1*(1), 16-22.

- Maharani, M., Ratnawati, D. E., & Rahayudi, B. (2021). Pengelompokan Kondisi Komputer pada Bank X di Seluruh Indonesia Menggunakan Metode K-Medoids Clustering. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN*, 2548, 964X.
- Melita, R. (2018). *Penerapan Metode Term Frequency Inverse Document Frequency (Tf-Idf) Dan Cosine Similarity Pada Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits Berbasis Web (Studi Kasus: Hadits Shahih Bukhari-Muslim)* (Bachelor's thesis, Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta).
- Mughnyanti, M., Efendi, S., & Zarlis, M. (2020). Analysis of determining centroid clustering x-means algorithm with davies-bouldin index evaluation. In *IOP Conference Series: Materials Science and Engineering* (Vol. 725, No. 1, p. 012128). IOP Publishing.
- Praja, B. S., Kusuma, P. D., & Setianingsih, C. (2019). Penerapan Metode K-Means Clustering Dalam Pengelompokan Data Penumpang Dan Kapal Angkutan Laut Di Indonesia. *eProceedings of Engineering*, 6(1).
- Pramesti, D. F., Furqon, M. T., & Dewi, C. (2017). Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Hotspot). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN*, 2548, 964X.
- Rizaldi, S. T., Al Khairi, A., & Mustakim, M. (2022). Text Mining Classification Opini Publik Terhadap Provider di Indonesia. In *Seminar Nasional Teknologi Informasi Komunikasi dan Industri* (pp. 79-86).
- Rofiqi, M. A., Fauzan, A. C., Agustin, A. P., & Saputra, A. A. (2019). Implementasi Term-Frequency Inverse Document Frequency (TF-IDF) Untuk Mencari Relevansi Dokumen Berdasarkan Query. *ILKOMNIKA: Journal of Computer Science and Applied Informatics*, 1(2), 58-64.
- Rosid, M. A., Fitriani, A. S., Astutik, I. R. I., Mulloh, N. I., & Gozali, H. A. (2020). Improving text preprocessing for student complaint document classification using sastrawi. In *IOP Conference Series: Materials Science and Engineering* (Vol. 874, No. 1, p. 012017). IOP Publishing.
- Salloum, S. A., Al-Emran, M., Monem, A. A., & Shaalan, K. (2018). Using text mining techniques for extracting information from research articles. In *Intelligent natural language processing: Trends and Applications* (pp. 373-397). Springer, Cham.

- Sepri, D., & Fimazid, Y. (2021). Pengelompokan Penyebaran Covid-19 di Kota Padang Menggunakan Algoritma K-Medoids. *Insearch: Information System Research Journal*, 1(02), 39-45.
- Soni, K. G., & Patel, A. (2017). Comparative Analysis of K-means and K-medoids Algorithm on IRIS Data. *International Journal of Computational Intelligence Research*, 13(5), 899-906.
- Soyusiawaty, D., Jones, A. H. S., & Lestariw, N. L. (2020). The stemming application on affixed javanese words by using nazief and adriani algorithm. In *IOP Conference Series: Materials Science and Engineering* (Vol. 771, No. 1, p. 012026). IOP Publishing.
- Sundari, S., Damanik, I. S., Windarto, A. P., Tambunan, H. S., Jalaluddin, J., & Wanto, A. (2019). Analisis K-Medoids Clustering Dalam Pengelompokan Data Imunisasi Campak Balita Di Indonesia. In *Prosiding Seminar Nasional Riset Information Science (SENARIS)* (Vol. 1, pp. 687-696).
- Wibawa, A. P., Miftahuddin, F., & Suyono, S. (2021). K-Medoids Clustering untuk Pembentukan Database Stopword Bahasa Jawa. *Ranah: Jurnal Kajian Bahasa*, 10(2), 261-269.