

***CLUSTERING MENGGUNAKAN METODE K-MEDOIDS
PADA ARTIKEL BERITA ONLINE BERBAHASA INDONESIA***

Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Strata-1 Pada
Jurusan Teknik Informatika



Oleh :

Isnania Tuljannah
NIM : 09021181924026

**Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2025**

HALAMAN PENGESAHAN

SKRIPSI

CLUSTERING MENGGUNAKAN METODE K-MEDOIDS
PADA ARTIKEL BERITA ONLINE BERBAHASA INDONESIA

Sebagai salah satu syarat untuk penyelesaian studi di
Program Studi S1 Teknik Informatika

Oleh:
ISNANIA TULJANNAH
09021181924026

Pembimbing 1 : **Novi Yusliani, S.Kom., M.T.**
NIP. 198211082012122001

Pembimbing 2 : **M. Naufal Rachmatullah, S.Kom., M.T.**
NIP. 199212012022031008

Mengetahui
Ketua Jurusan Teknik Informatika



Hadipurnawan Satria, Ph.D
198004182020121001

TANDA LULUS UJIAN KOMPREHENSIF

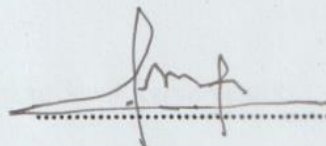
Pada hari Jum'at tanggal 25 Juli 2025 telah dilaksanakan Ujian Komprehensif Skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Isnania Tuljannah
NIM : 09021181924026
Judul : *Clustering Menggunakan Metode K-Medoids Pada Artikel Berita Online Berbahasa Indonesia*

dan dinyatakan **LULUS**.

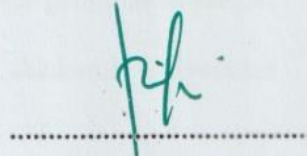
1. Ketua Penguji

Prof. Dr. Ermatita, M.Kom.
NIP. 196709132006042001



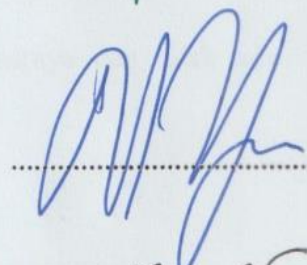
2. Penguji

Rifkie Primartha, S.T., M.T.
NIP. 197706012009121004



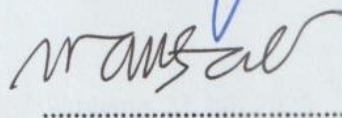
3. Pembimbing I

Novi Yusliani, S.Kom., M.T.
NIP. 198211082012122001



4. Pembimbing II

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



Mengetahui,
Ketua Jurusan Teknik Informatika



Hadipurnawan Satria, Ph.D.
NIP. 198004182020121001

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah in:

Nama : Isnania Tuljannah
NIM : 09021181924026
Program Studi : Teknik Informatika
Judul Skripsi : *Clustering Menggunakan Metode K-Medoids* Pada Artikel
Berita *Online* Berbahasa Indonesia

Hasil Pengecekan *Software iThenticate/Turnitin*: 5%

Menyatakan bahwa laporan penelitian saya merupakan hasil karya sendiri dan 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 dan tidak ada paksaan oleh siapapun.



Palembang, 21 Juli 2025



Isnania Tuljannah
NIM. 09021181924026

MOTTO DAN PERSEMBAHAN

“Kesuksesan bukan milik orang pintar, tetapi milik mereka yang senantiasa berusaha.”

- B.J. Habibie

“Langkahku memang tak cepat, tapi tak pernah benar-benar berhenti.”

- Isnania Tuljannah

Saya persembahkan karya tulis ini kepada:

- Kedua Orang Tua
- Adik
- Keluarga Besar
- Dosen Pembimbing
- Jurusan Teknik Informatika
- Fakultas Ilmu Komputer
- Universitas Sriwijaya

CLUSTERING USING K-MEDOIDS METHOD ON INDONESIAN ONLINE NEWS ARTICLES

By :

Isnania Tuljannah
09021181924026

ABSTRACT

The rapid growth of Indonesian online news articles presents challenges in clustering based on content similarity. This study aims to develop a clustering system for online news articles using the k-medoids method. A total of 500 articles from Kaggle were processed through text preprocessing (case folding, tokenizing, stopword removal, and stemming), TF-IDF weighting, and clustering with k-medoids. Evaluation using the davies bouldin index (DBI) showed the optimal result at 10 clusters with a DBI value of 8.2918. Each cluster represented specific themes such as health, politics, economy, entertainment, culture, and tourism. Wordcloud visualization and word frequency analysis strengthened topic interpretation. The findings demonstrate that k-medoids is effective in clustering Indonesian online news articles and supports text-based information analysis.

Keywords: *Clustering, K-Medoids, TF-IDF, Online News Articles, Davies Bouldin Index (DBI)*

CLUSTERING MENGGUNAKAN METODE *K-MEDOIDS* PADA ARTIKEL BERITA *ONLINE* BERBAHASA INDONESIA

Oleh :

Isnania Tuljannah
09021181924026

ABSTRAK

Perkembangan pesat artikel berita *online* berbahasa Indonesia menimbulkan tantangan dalam pengelompokan berita berdasarkan kesamaan isi. Penelitian ini bertujuan mengembangkan sistem *clustering* artikel berita *online* menggunakan metode *k-medoids*. Data berupa 500 artikel dari Kaggle diproses melalui tahapan *text preprocessing* (*case folding*, *tokenizing*, *stopword removal*, *stemming*), pembobotan TF-IDF, dan *clustering* dengan *k-medoids*. Evaluasi menggunakan *davies bouldin index* (DBI) menunjukkan hasil optimal pada 10 *cluster* dengan nilai DBI 8.2918. Setiap *cluster* merepresentasikan tema tertentu, seperti kesehatan, politik, ekonomi, hiburan, budaya, dan pariwisata. Visualisasi *wordcloud* dan frekuensi kata memperkuat interpretasi topik. Hasil penelitian membuktikan *k-medoids* efektif dalam mengelompokkan artikel berita *online* serta mendukung analisis informasi berbasis teks.

Kata Kunci: *Clustering*, *K-Medoids*, TF-IDF, Artikel Berita *Online*, *Davies Bouldin Index* (DBI)

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan penyusunan skripsi yang berjudul “*Clustering Menggunakan Metode K-Medoids Pada Artikel Berita Online Berbahasa Indonesia*”. Skripsi ini disusun sebagai salah satu syarat dalam menyelesaikan Pendidikan Strata-1 Jurusan Teknik Informatika, Fakultas Ilmu Komputer, Universitas Sriwijaya.

Perjalanan panjang dalam menyelesaikan pendidikan ini tentu tidak mudah. Banyak hal yang telah dilalui, baik tantangan, keterlambatan, maupun keraguan. Namun semua itu menjadi bagian dari proses pembelajaran dan pendewasaan diri. Dengan rasa penuh syukur, akhirnya skripsi ini dapat diselesaikan meskipun tidak dalam waktu yang cepat. Dalam proses penyusunan skripsi ini, penulis banyak mendapatkan bantuan, bimbingan, serta dukungan dari berbagai pihak. Oleh karena itu, dengan segala kerendahan hati, penulis ingin menyampaikan terima kasih kepada:

1. Bapak Prof. Dr. Erwin, S.Si., M.Si. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
2. Bapak Hadipurnawan Satria, Ph.D. selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
3. Ibu Novi Yusliani, S.Kom., M.T selaku Dosen Pembimbing I dan Dosen Pembimbing Akademik yang telah memberikan bimbingan, arahan, dan motivasi dengan penuh kesabaran hingga skripsi ini dapat terselesaikan.
4. Bapak M. Naufal Rachmatullah, S.Kom., M.T. selaku Dosen Pembimbing

II yang telah membantu memberikan arahan dan bimbingan dan juga motivasi selama proses penyusunan skripsi ini.

5. Ibu Prof. Dr. Ermatita, M.Kom. selaku Ketua Penguji, serta Bapak Rifkie Primartha, S.T., M.T. selaku Penguji, yang telah memberikan saran dan masukan yang sangat berarti dalam penyempurnaan penelitian ini.
6. Seluruh Dosen Jurusan Teknik Informatika yang telah memberikan ilmu dan pengalaman selama masa perkuliahan.
7. Seluruh Staf dan Pegawai Fakultas Ilmu Komputer yang telah membantu kelancaran proses administrasi dan akademik selama masa perkuliahan.
8. Kedua orang tua tercinta, yang tidak pernah lelah mendo'akan, memberi semangat, serta dukungan baik secara moril maupun materiil di setiap langkah perjalanan panjang ini. Terima kasih atas kesabaran dan cinta yang tak ternilai, yang menjadi fondasi utama dalam menyelesaikan pendidikan ini hingga akhir.
9. Adik tercinta, yang selalu menjadi sumber semangat dan pengingat untuk terus berusaha dan tidak menyerah.
10. Keluarga besar yang telah memberikan do'a dan dukungan selama masa perkuliahan dan penyusunan skripsi ini. Terima kasih atas perhatian yang menjadi penyemangat tersendiri bagi penulis.
11. Teman-teman seperjuangan yang telah lulus terlebih dahulu, atas inspirasi dan motivasi yang tanpa disadari turut mendorong penulis untuk bangkit dan menyelesaikan skripsi ini.
12. Untuk teman seperjuangan penulis, terima kasih telah menjadi tempat

berbagi keluh kesah, tawa, dan semangat, serta menjadi pengingat untuk terus melangkah dan tidak menyerah.

13. Semua pihak yang tidak dapat penulis sebutkan satu persatu, namun kontribusinya sangat berarti.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih terdapat banyak kekurangan. Oleh karena itu, kritik dan saran yang membangun sangat penulis harapkan. Semoga skripsi ini dapat memberikan manfaat bagi pembaca dan menjadi referensi penelitian selanjutnya.

Palembang, 20 Juli 2025

Isnania Tuljannah
NIM. 09021181924026

DAFTAR ISI

Halaman

HALAMAN JUDUL.....	i
HALAMAN 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.....	xi
DAFTAR TABEL	xiv
DAFTAR GAMBAR	xv

BAB I PENDAHULUAN	I-1
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 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 Preprocessing</i>	II-4
2.2.3 Pembobotan Kata	II-7
2.2.4 <i>K-Medoids</i>	II-10
2.2.5 <i>Davies Bouldin Index</i>	II-12
2.2.6 Berita.....	II-14
2.2.7 <i>Rational Unified Process</i>	II-15
2.3 Penelitian Lain yang Relevan.....	II-17
2.4 Kesimpulan	II-18

BAB III METODOLOGI PENELITIAN.....	III-1
3.1 Pendahuluan	III-1
3.2 Pengumpulan Data	III-1
3.2.1 Jenis dan Sumber Data.....	III-1
3.2.2 Metode Pengumpulan Data.....	III-1
3.3 Tahapan Penelitian	III-1
3.3.1 Mengumpulkan Data.....	III-3
3.3.2 Menentukan Kerangka Kerja Penelitian	III-4
3.3.3 Menentukan Kriteria Pengujian	III-6
3.3.4 Menetapkan Format Data Pengujian.....	III-6
3.3.5 Menentukan Alat Bantu Penelitian	III-7
3.3.6 Melakukan Pengujian Penelitian	III-8
3.3.7 Melakukan Analisis Hasil Pengujian dan Membuat Kesimpulan Penelitian	III-8
3.4 Metode Pengembangan Perangkat Lunak	III-8
3.4.1 Tahap Insepsi	III-8
3.4.2 Tahap Elaborasi	III-9
3.4.3 Tahap Konstruksi.....	III-9
3.4.4 Tahap Transisi.....	III-10
3.5 Manajemen Proyek Penelitian.....	III-11
3.6 Kesimpulan.....	III-12
 BAB IV PENGEMBANGAN PERANGKAT LUNAK	IV-1
4.1 Pendahuluan	IV-1
4.2 Tahap Insepsi.....	IV-1
4.2.1 Pemodelan Bisnis.....	IV-1
4.2.2 Analisis Kebutuhan Sistem.....	IV-2
4.2.3 Analisis Desain	IV-3
4.2.3.1 Analisis Perangkat Lunak.....	IV-3
4.2.3.2 Analisis Data.....	IV-4
4.2.3.3 Analisis Metode <i>K-Medoids</i>	IV-5
4.2.3.4 Desain Perangkat Lunak	IV-14
4.3 Tahap Elaborasi.....	IV-18
4.3.1 Pemodelan Bisnis.....	IV-18
4.3.1.1 Perancangan Data	IV-19
4.3.1.2 Perancangan Antarmuka	IV-20
4.3.2 Kebutuhan Sistem	IV-22
4.3.3 Diagram	IV-23
4.3.3.1 Diagram Aktivitas.....	IV-23
4.3.3.2 Diagram Alur	IV-25
4.4 Tahap Konstruksi	IV-26
4.4.1 Kebutuhan Sistem	IV-26
4.4.2 Diagram Kelas	IV-28
4.4.3 Implementasi.....	IV-28
4.4.3.1 Implementasi Kelas	IV-29

4.4.3.2 Implementasi Antarmuka.....	IV-30
4.5 Tahap Transisi	IV-32
4.5.1 Pemodelan Bisnis.....	IV-32
4.5.2 Implementasi.....	IV-32
4.6 Kesimpulan.....	IV-34
 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-2
5.3 Analisis Hasil Pengujian	V-4
5.4 Kesimpulan.....	V-21
 BAB VI KESIMPULAN DAN SARARAN.....	 VI-1
6.1 Kesimpulan.....	VI-1
6.2 Saran.....	VI-2
 DAFTAR PUSTAKA	 xvi
LAMPIRAN.....	xix

DAFTAR TABEL

Halaman

Tabel III-1 Sampel Data Artikel Berita <i>Online</i>	III-3
Tabel III-2 Rancangan Tabel Hasil Uji <i>Clustering</i>	III-7
Tabel III-3 Rencana Manajemen Proyek Penelitian	III-11
Tabel IV-1 Kebutuhan Fungsional.....	IV-2
Tabel IV-2 Kebutuhan NonFungsional.....	IV-3
Tabel IV-3 Contoh Dokumen.....	IV-6
Tabel IV-4 Proses <i>Case Folding</i>	IV-6
Tabel IV-5 Proses <i>Tokenizing</i>	IV-7
Tabel IV-6 Proses <i>Stopword Removal</i>	IV-7
Tabel IV-7 Proses <i>Stemming</i>	IV-7
Tabel IV-8 Hasil <i>Preprocessing</i> Artikel Berita <i>Online</i>	IV-8
Tabel IV-9 Proses Perhitungan Pembobotan Kata.....	IV-8
Tabel IV-10 Hasil <i>Clustering</i>	IV-14
Tabel IV-11 Definisi Aktor.....	IV-15
Tabel IV-12 Definisi <i>Use Case</i>	IV-15
Tabel IV-13 Skenario <i>Use Case</i>	IV-16
Tabel IV-14 Implementasi Kelas	IV-29
Tabel IV-15 Pengujian <i>Use Case</i> Melakukan <i>Clustering</i>	IV-33
Tabel V-1 Hasil Uji <i>Clustering</i>	V-3
Tabel V-2 <i>Frekuensi</i> Kata pada <i>Cluster</i> 1.....	V-6
Tabel V-3 <i>Frekuensi</i> Kata pada <i>Cluster</i> 2.....	V-7
Tabel V-4 <i>Frekuensi</i> Kata pada <i>Cluster</i> 3.....	V-9
Tabel V-5 <i>Frekuensi</i> Kata pada <i>Cluster</i> 4.....	V-10
Tabel V-6 <i>Frekuensi</i> Kata pada <i>Cluster</i> 5.....	V-11
Tabel V-7 <i>Frekuensi</i> Kata pada <i>Cluster</i> 6.....	V-13
Tabel V-8 <i>Frekuensi</i> Kata pada <i>Cluster</i> 7.....	V-14
Tabel V-9 <i>Frekuensi</i> Kata pada <i>Cluster</i> 8.....	V-15
Tabel V-10 <i>Frekuensi</i> Kata pada <i>Cluster</i> 9.....	V-17
Tabel V-11 <i>Frekuensi</i> Kata pada <i>Cluster</i> 10.....	V-18

DAFTAR GAMBAR

Halaman

Gambar II-1 Alur Proses <i>Clustering</i>	II-3
Gambar II-2 <i>Case Folding</i>	II-4
Gambar II-3 <i>Tokenizing</i>	II-5
Gambar II-4 <i>Stopword Removal</i>	II-6
Gambar II-5 <i>Stemming</i>	II-7
Gambar II-6 Ilustrasi Metode TF-IDF	II-8
Gambar II-7 Arsitektur Metode RUP.....	II-16
Gambar III-1 Tahapan Penelitian.....	III-2
Gambar III-2 Kerangka Kerja Penelitian.....	III-5
Gambar IV-1 Diagram <i>Use Case</i>	IV-14
Gambar IV-2 Rancangan Antarmuka Halaman Utama Sistem.....	IV-20
Gambar IV-3 Rancangan Antarmuka Halaman Hasil <i>Clustering</i>	IV-21
Gambar IV-4 Diagram Aktivitas Melakukan <i>Clustering</i>	IV-24
Gambar IV-5 Diagram Alur Melakukan <i>Clustering</i>	IV-25
Gambar IV-6 Diagram Kelas	IV-28
Gambar IV-7 Implementasi Antarmuka Halaman Utama Sistem.....	IV-30
Gambar IV-8 Implementasi Antarmuka Halaman Hasil <i>Clustering</i>	IV-31
Gambar V-1 <i>Wordcloud</i> pada <i>Cluster 1</i>	V-6
Gambar V-2 <i>Wordcloud</i> pada <i>Cluster 2</i>	V-7
Gambar V-3 <i>Wordcloud</i> pada <i>Cluster 3</i>	V-8
Gambar V-4 <i>Wordcloud</i> pada <i>Cluster 4</i>	V-10
Gambar V-5 <i>Wordcloud</i> pada <i>Cluster 5</i>	V-11
Gambar V-6 <i>Wordcloud</i> pada <i>Cluster 6</i>	V-12
Gambar V-7 <i>Wordcloud</i> pada <i>Cluster 7</i>	V-14
Gambar V-8 <i>Wordcloud</i> pada <i>Cluster 8</i>	V-15
Gambar V-9 <i>Wordcloud</i> pada <i>Cluster 9</i>	V-16
Gambar V-10 <i>Wordcloud</i> pada <i>Cluster 10</i>	V-18

BAB I

PENDAHULUAN

1.1 Pendahuluan

Pada bab ini membahas mengenai pokok pikiran yang akan dibahas dalam penelitian ini meliputi latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, sistematika penulisan, dan kesimpulan.

1.2 Latar Belakang Masalah

Berita merupakan salah satu kebutuhan utama bagi setiap orang di berbagai belahan dunia. Melalui berita, seseorang dapat memperoleh berbagai informasi terkini tentang berbagai peristiwa dan isu yang sedang terjadi (Riani et al., 2018). Berita sendiri diartikan sebagai informasi mengenai kejadian atau pendapat yang penting dan menarik yang dipublikasikan secara luas di media massa (Frinta et al., 2019). Seiring meningkatnya kebutuhan masyarakat akan informasi, media berita pun mengalami perkembangan yang pesat, terutama dalam bentuk media daring atau berita *online*. Berita *online* memiliki keunggulan dalam hal jangkauan pembaca yang luas dan efisiensi waktu akses, sehingga semakin diminati oleh pengguna internet (Frinta et al., 2019).

Situs berita *online* menyajikan berbagai jenis artikel dengan topik yang sangat beragam, seperti politik, ekonomi, kesehatan, kriminal, hingga bencana alam. Namun demikian, pengelompokan artikel-artikel tersebut ke dalam kategori tertentu sebagian besar masih dilakukan secara manual. Proses pengelompokan secara manual ini memerlukan waktu yang lama dan tenaga yang besar karena

harus membaca dan memahami isi berita terlebih dahulu agar didapat pengelompokan yang tepat. Hal ini menyebabkan kurang optimalnya pemanfaatan informasi yang tersedia dalam jumlah besar (Riani et al., 2018).

Jumlah dokumen digital seperti artikel berita yang terus bertambah setiap hari menjadikannya sebagai sumber data yang sangat berharga. Informasi ini dapat digunakan untuk berbagai keperluan, seperti analisis opini publik, deteksi isu sosial, hingga pengambilan keputusan berbasis data. Namun, besarnya volume data dengan dimensi tinggi dan konten yang tidak terstruktur menjadi tantangan dalam proses pengelolaan dan pencarian informasi. Terlebih lagi, sebagian besar artikel berita tidak disertai dengan label, sehingga sulit untuk dianalisis secara manual (Simanjuntak et al., 2023).

Untuk mengatasi permasalahan tersebut, pendekatan *data mining* khususnya metode *clustering* menjadi salah satu solusi yang tepat. *Clustering* adalah teknik *unsupervised learning* yang dapat digunakan untuk mengelompokkan data tidak berlabel ke dalam beberapa kelompok berdasarkan kemiripan fitur (Simanjuntak et al., 2023). Dalam konteks berita *online*, teknik ini dikenal sebagai *text clustering*, yaitu proses mengelompokkan dokumen teks ke dalam kategori berdasarkan kesamaan isi. Dengan penerapan *clustering*, artikel-artikel berita yang serupa dapat dikelompokkan secara otomatis, sehingga memudahkan pengguna dalam menemukan informasi yang relevan serta mempercepat proses analisis isi dokumen (Wardy et al., 2022).

Salah satu algoritma yang dapat digunakan untuk melakukan *clustering* adalah *k-medoids*. Algoritma ini merupakan varian dari *k-means* yang dikenal

lebih tahan terhadap *noise* dan *outlier* dalam data (Asmiatun et al., 2020). *K-medoids* menggunakan objek aktual sebagai pusat *cluster* (*medoid*), bukan nilai rata-rata seperti pada *k-means*. Dengan demikian, hasil pengelompokan yang dihasilkan oleh *k-medoids* lebih stabil dan representatif, terutama pada data teks berdimensi tinggi seperti artikel berita. Algoritma ini juga cocok digunakan untuk metode partisi, yaitu membagi data ke dalam sejumlah *cluster* berdasarkan kedekatan antar data (Gunawan et al., 2020).

Beberapa penelitian mengenai *clustering* telah dilakukan sebelumnya. Seperti penelitian (Wardy et al., 2022) mengelompokkan artikel berita pada portal berita *online* menggunakan metode *k-means*. Hasil evaluasi menunjukkan bahwa metode *k-means* mampu menghasilkan performa *clustering* yang tinggi, dengan nilai *accuracy* sebesar 94,93%, *precision* sebesar 95,07%, *recall* sebesar 94,93%, dan *f1-score* sebesar 94,94%. Namun, *k-means* sangat sensitif terhadap *outlier* karena menggunakan nilai rata-rata (*mean*) sebagai pusat *cluster*. Keberadaan *outlier* dapat mempengaruhi posisi *centroid* dan menyebabkan hasil *clustering* menjadi kurang representatif.

Penelitian (Ramadhani et al., 2022) membandingkan algoritma *k-means* dan *k-medoids* dalam mengelompokkan dokumen tesis. Hasil penelitian tersebut menunjukkan bahwa algoritma *k-medoid* memiliki kinerja yang lebih baik dibandingkan dengan *k-means*. Hal tersebut ditunjukkan dengan hasil rata-rata nilai DBI yang lebih rendah pada algoritma *k-medoids*, yaitu sebesar 1,56, dibandingkan dengan algoritma *k-means* yang memiliki rata-rata nilai DBI sebesar 2,79.

Penelitian (Sepyanto et al., 2020) menggunakan algoritma *k-medoids* untuk mengelompokkan data *tweet* masyarakat terhadap acara Indonesia Lawyers Club. Hasil penelitian tersebut menunjukkan bahwa metode *k-medoids* berhasil diterapkan dalam proses pengelompokan data. Evaluasi hasil *clustering* dilakukan menggunakan metode *silhouette coefficient* dan menghasilkan nilai sebesar 0,19 dan mampu memprediksi opini masyarakat melalui media sosial twitter dengan tingkat akurasi mencapai 80%.

Penelitian ini bertujuan untuk mengembangkan sistem yang dapat mengelompokkan artikel berita *online* berbahasa Indonesia secara otomatis menggunakan metode *k-medoids*.

1.3 Rumusan Masalah

Rumusan masalah dari penelitian ini adalah sebagai berikut:

1. Bagaimana mengembangkan sistem *clustering* artikel berita *online* menggunakan metode *k-medoids*?
2. Bagaimana kinerja metode *k-medoids* dalam melakukan *clustering* artikel berita *online* berdasarkan *davies bouldin index* (DBI)?

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Menghasilkan sistem *clustering* artikel berita *online* menggunakan metode *k-medoids*.
2. Mengetahui kinerja metode *k-medoids* dalam melakukan *clustering* artikel berita *online* berdasarkan *davies bouldin index* (DBI).

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut:

1. Mempermudah dalam proses pengelompokan artikel berita *online* tanpa harus membacanya satu persatu.
2. Penelitian ini dapat menjadi rujukan untuk penelitian di masa mendatang.

1.6 Batasan Masalah

Batasan masalah dari penelitian ini adalah sebagai berikut:

1. Data yang digunakan dalam penelitian ini berupa artikel berita *online* dengan ekstensi *.csv*.
2. Data yang digunakan dalam penelitian ini diambil dari situs Kaggle.

1.7 Sistematika Penulisan

Sistematika penulisan skripsi ini adalah sebagai berikut:

BAB I PENDAHULUAN

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

BAB II KAJIAN LITERATUR

Pada bab ini membahas mengenai landasan teori yang digunakan dalam penelitian, seperti *clustering*, *text preprocessing*, pembobotan kata, algoritma *k-medoids*, metode evaluasi *clustering*, artikel berita *online*, metode pengembangan perangkat lunak, dan penelitian terdahulu yang relevan terhadap penelitian ini.

BAB III METODOLOGI PENELITIAN

Pada bab ini membahas mengenai metode penelitian, tahapan yang akan

dilaksanakan pada penelitian, dan metode pengembangan perangkat lunak. Tahapan penelitian akan dibahas secara rinci berdasarkan kerangka kerja tertentu. Di bagian akhir bab ini terdapat perancangan manajemen proyek pada penelitian ini.

BAB IV PENGEMBANGAN PERANGKAT LUNAK

Pada bab ini membahas mengenai proses pengembangan perangkat lunak *clustering* artikel berita *online* menggunakan metode *k-medoids*. Metode pengembangan perangkat lunak yang diterapkan dalam penelitian ini berorientasi objek yaitu dengan menggunakan metode *Rational Unified Process* (RUP). Tahapan-tahapan RUP meliputi tahap insepisi, elaborasi, konstruksi, dan transisi.

BAB V HASIL DAN ANALISIS PENELITIAN

Pada bab ini membahas dan menganalisis mengenai hasil pengujian yang telah dilakukan berdasarkan langkah-langkah yang telah direncanakan.

BAB VI KESIMPULAN DAN SARAN

Pada bab ini membahas mengenai kesimpulan dari penelitian yang telah dilakukan, serta saran-saran yang dapat dijadikan acuan untuk pengembangan dan penyempurnaan penelitian di masa mendatang.

1.8 Kesimpulan

Berdasarkan uraian sebelumnya, maka akan dikembangkan sistem *clustering* artikel berita *online* menggunakan metode *k-medoids*. Metode *k-medoids* diharapkan dapat diimplementasikan dengan baik untuk mengembangkan sistem *clustering* artikel berita *online* dalam penelitian ini.

DAFTAR PUSTAKA

- Alamtaha, Z., Djakaria, I., & Yahya, N. I. (2023). Implementasi Algoritma Hierarchical Clustering dan Non-Hierarchical Clustering untuk Pengelompokan Pengguna Media Sosial. *Journal of Statistics and Its Application*, 4(1), 33–34. <https://doi.org/10.20956/ejsa.vi.24830>
- Asmiatun, S., Wakhidah, N., & Putri, A. N. (2020). Penerapan Metode K-Medoids Untuk Pengelompokan Kondisi Jalan Di Kota Semarang. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 6(2), 171–180. <https://doi.org/10.35957/jatisi.v6i2.193>
- Fialine, A. P., Alodia, D. A., Endriani, D., & Widodo, E. (2021). Implementasi Metode K-Medoids Clustering untuk Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Pendidikan. *Sepren*, 2(2), 1–13. <https://doi.org/10.36655/sepren.v2i2.606>
- Fitrayana, P. R., & Saputro, D. R. S. (2022). Algoritme Clustering Large Application (CLARA) untuk Menangani Data Outlier. *PRISMA, Prosiding Seminar Nasional Matematika*, 5, 721–725.
- Frinta, K., Indriati, & Adikara, P. P. (2019). Pencarian Berita Berbahasa Indonesia Menggunakan Metode BM25. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 3(3), 2589–2595. <http://j-ptiik.ub.ac.id>
- Gunawan, I., Anggraeni, G., Rini, S., Putri, Y. M., & Zikri, Y. K. (2020). Klasterisasi provinsi di Indonesia berbasis perkembangan kasus Covid-19 menggunakan metode K-Medoids. *Seminar Nasional Matematika Dan Pendidikan Matematika (5thSENATIK)*, 301–306.
- Herwijayanti, B., Ratnawati, D. E., & Muflikhah, L. (2018). Klasifikasi Berita Online dengan menggunakan Pembobotan TF-IDF dan Cosine Similarity. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(1), 306–312. <http://j-ptiik.ub.ac.id>
- Kurniadevy. (n.d.). *berita* [Data set]. Kaggle. Retrieved November 2024, from <https://www.kaggle.com/datasets/kurniadevy/berita>
- Nabila, Z., Isnain, A. R., Permata, & Abidin, Z. (2021). Analisis data mining untuk clustering kasus Covid-19 di Provinsi Lampung dengan algoritma K-Means. *Jurnal Teknologi dan Sistem Informasi (JTSI)*, 2(2), 100–108. <http://jim.teknokrat.ac.id/index.php/JTSI>
- Ningsih, S., Suhada, S., Dewi, R., & Windarto, A. P. (2019). Pengklasteran

Dokumen Tentang Dispensasi Nikah Menggunakan Algoritma K-Medoids. *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, 1(September), 677. <https://doi.org/10.30645/senaris.v1i0.74>

Perwitasari, R., Afawani, R., & Anjarwani, S. E. (2020). Penerapan Metode Rational Unified Process (RUP) Dalam Pengembangan Sistem Informasi Medical Check Up Pada Citra Medical Centre. *Jurnal Teknologi Informasi, Komputer, Dan Aplikasinya (JTIKA)*, 2(1), 76–88. <https://doi.org/10.29303/jtika.v2i1.85>

Purniawan, I. M. A. B. M. A. T.-I. D. K.-M. D. M. S. D. C. P. S. D. C., Sasmita, G. M. A., & Pratama, I. P. A. E. (2022). Clustering Berita Menggunakan Algoritma Tf-Idf Dan K-Means Dengan Memanfaatkan Sumber Data Crawling Pada Situs Detik.Com. *JITTER- Jurnal Ilmiah Teknologi Dan Komputer*, 3(1), 821–830.

Putra, B. J. M., & Yuniarti, D. A. F. (2020). Analisis Hasil Belajar Mahasiswa Dengan Clustering Menggunakan Metode K-Means. *Poros Teknik*, 12(2), 49–58. <https://ejurnal.poliban.ac.id/index.php/porosteknik/article/view/927>

Rahmah, A. S. (2020). Klasterisasi Pola Penjualan Pestisida Menggunakan Metode K-Means Clustering (Studi Kasus Di Toko Juanda Tani Kecamatan Hutabayu Raja). *Djtechno: Jurnal Teknologi Informasi*, 1(1), 1–5. <https://doi.org/10.46576/djtechno.v1i1.964>

Rahmawati, L., Widya Sihwi, S., & Suryani, E. (2016). Analisa Clustering Menggunakan Metode K-Means Dan Hierarchical Clustering (Studi Kasus : Dokumen Skripsi Jurusan Kimia, Fmipa, Universitas Sebelas Maret). *Jurnal Teknologi & Informasi ITSmart*, 3(2), 66. <https://doi.org/10.20961/its.v3i2.654>

Ramadhani, S., Azzahra, D., & Z, T. (2022). Comparison of K-Means and K-Medoids Algorithms in Text Mining based on Davies Bouldin Index Testing for Classification of Student's Thesis. *Digital Zone: Jurnal Teknologi Informasi Dan Komunikasi*, 13(1), 24–33. <https://doi.org/10.31849/digitalzone.v13i1.9292>

Randhika, M. N., Young, J. C., Suryadibrata, A., & Mandala, H. (2021). Implementasi Algoritma Complement dan Multinomial Naïve Bayes Classifier Pada Klasifikasi Kategori Berita Media Online. *Ultimatics : Jurnal Teknik Informatika*, 13(1), 19–25. <https://doi.org/10.31937/ti.v13i1.1921>

Rani, G. P. I., Aziz, A., & T.S, M. P. (2022). Implementasi Euclidean Dan Chebyshev Distance Pada K-Medoids Clustering. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 6(2), 710–715. <https://doi.org/10.36040/jati.v6i2.5443>

Riani, Hamzah, A., & N, E. K. (2018). *Jurnal SCRIPT Vol . 6 No . 1 Desember 2018 ISSN : 2338-6304 Jurnal SCRIPT Vol . 6 No . 1 Desember 2018*. 6(1),

- Riyanto, R. Y., Mawardi, V. C., & Perdana, N. J. (2023). Clustering Berita Sepak Bola Dengan Metode K-Means. *Jurnal Ilmu Komputer Dan Sistem Informasi*, 11(1). <https://doi.org/10.24912/jiksi.v11i1.24072>
- Rofiqi, A. Y. (2017). Clustering Berita Olahraga Berbahasa Indonesia Menggunakan Metode K-Medoid Bersyarat. *Jurnal Simantec*, 6(1), 25–32.
- Roihan, A., Sunarya, P. A., & Rafika, A. S. (2020). *Pemanfaatan Machine Learning dalam Berbagai Bidang : Review paper*. 5(April), 75–82.
- Runturamby, E. M. W., Rantung, V. P., & Santa, K. (2024). *Peringkasan Teks Otomatis Berita Online Komisi Pemilihan Umum Menggunakan Algoritma K-Means Clustering*. 299–312.
- Sepyanto, K. J. S., Chrisnanto, Y. H., & Umbara, F. R. (2020). Sistem Segmentasi Program Talk Show Berdasarkan Media Sosial Twitter Menggunakan Metode K-Medoids Clustering. *Prosiding SISFOTEK*, 342–347. <http://seminar.iaii.or.id/index.php/SISFOTEK/article/view/243>
- Simanjuntak, H. T. A., Silaban, P. E. P., Manurung, J. K. S., & Sormin, V. H. (2023). Klasterisasi Berita Bahasa Indonesia Dengan Menggunakan K-Means Dan Word Embedding. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 10(3), 641–652. <https://doi.org/10.25126/jtiik.20231026468>
- Sulistyawati, A. A. D., & Sadikin, M. (2021). Penerapan algoritma K-Medoids untuk menentukan segmentasi pelanggan. *SISTEMASI: Jurnal Sistem Informasi*, 10(3), 516–526. <http://sistemasi.ftik.unisi.ac.id>
- Utami, N. W., & Putra, I. G. J. E. (2022). Text Minig Clustering Untuk Pengelompokan Topik Dokumen Penelitian Menggunakan Algoritma K-Means Dengan Cosine Similarity. *Jurnal Informatika Teknologi Dan Sains*, 4(3), 255–259. <https://doi.org/10.51401/jinteks.v4i3.1907>
- Wardy, D. K., Putra, I. K. G. D., & Rusjyanthi, N. K. D. (2022). Clustering Artikel pada Portal Berita Online. *JITTER- Jurnal Ilmiah Teknologi Dan Komputer*, 3(1), 3–11.
- Widyanto, A. (2020). Penerapan Metode RUP pada Sistem Informasi Unit Kegiatan Mahasiswa STMIK PalComTech. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 9(3), 323–331. <https://doi.org/10.32736/sisfokom.v9i3.789>