PERINGKASAN TEKS EKSTRAKTIF KARYA ILMIAH MAHASISWA MENGGUNAKAN *FUZZY C-MEANS* DAN *VECTOR SPACE MODEL*

SKRIPSI

Oleh :
VIVIN OCTAVIA CAHYANI
NIM. 210605110038

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2025

PERINGKASAN TEKS EKSTRAKTIF KARYA ILMIAH MAHASISWA MENGGUNAKAN *FUZZY C-MEANS* DAN *VECTOR SPACE MODEL*

SKRIPSI

Diajukan kepada:
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh :
VIVIN OCTAVIA CAHYANI
NIM. 210605110038

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2025

HALAMAN PERSETUJUAN

PERINGKASAN TEKS EKSTRAKTIF KARYA ILMIAH MAHASISWA MENGGUNAKAN *FUZZY C-MEANS* DAN *VECTOR SPACE MODEL*

SKRIPSI

Oleh :
VIVIN OCTAVIA CAHYANI
NIM. 210605110038

Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal: 02 Desember 2024

Pembimbing I,	Pembimbing II,
Dr. Muhammad Faisal, M.T	Dr. Irwan Budi Santoso, M.Kom
NIP. 19740510 200501 1 007	NIP. 19770103 201101 1 004

Mengetahui,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Dr. Ir. Fachrul Kurniawan, M.MT, IPU
NIP. 19771020 200912 1 001

HALAMAN PENGESAHAN

PERINGKASAN TEKS EKSTRAKTIF KARYA ILMIAH MAHASISWA MENGGUNAKAN *FUZZY C-MEANS* DAN *VECTOR SPACE MODEL*

SKRIPSI

Oleh :
VIVIN OCTAVIA CAHYANI
NIM. 210605110038

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal: 06 Desember 2024

Susunan Dewan Penguji

Ketua Penguji	: Hani Nurhayati, M.T NIP. 19780625 200801 2 006	( )
Anggota Penguji I	: Tri Mukti Lestari, M.Kom NIP. 19911108 202012 2 005	( )
Anggota Penguji II	: Dr. Muhammad Faisal, M.T NIP. 19740510 200501 1 007	( )
Anggota Penguji III	: Dr. Irwan Budi Santoso, M.Kom NIP. 19770103 201101 1 004	( )

Mengetahui dan Mengesahkan,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Dr. Ir. Fachrul Kurniawan, M.MT, IPU
NIP. 19771020 200912 1 001

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama	: Vivin Octavia Cahyani
NIM	: 210605110038
Fakultas / Program Studi	: Sains dan Teknologi / Teknik Informatika
Judul Skripsi	: Peringkasan Teks Ekstraktif Karya Ilmiah Mahasiswa Menggunakan Fuzzy C-Means dan Vector Space Model

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 17 April 2025
Yang membuat pernyataan,

Vivin Octavia Cahyani
NIM.210605110038

HALAMAN MOTTO

“All is well”

HALAMAN PERSEMBAHAN

Puji syukur atas kehadirat Allah Subhanahu wa ta'ala, karena berkat rahmat dan petunjuk-Nya penulis dapat menyelesaikan skripsi ini.

Penulis ingin mempersembahkan skripsi ini kepada kedua kedua orang tua, keluarga, dosen, guru, sahabat, dan semua pihak yang telah membantu secara aktif dalam menyelesaikan penelitian ini.

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Puji syukur penulis panjatkan kepada Allah SWT yang senantiasa memberikan rahmat dan kesehatan, sehingga penulis mampu menyelesaikan skripsi ini dengan baik. Penulis menyampaikan ucapan Terimakasih kepada semua pihak yang pernah terlibat langsung maupun tidak langsung dalam menyelesaikan skripsi ini, bukan hanya karena usaha keras dari penulis sendiri, akan tetapi karena adanya dukungan dari berbagai pihak. Oleh karena itu penulis berterima kasih kepada:

1. Prof. Dr. M. Zainuddin, M.A., selaku rektor Universitas islam Negeri Maulana Malik Ibrahim Malang.
2. Prof. Dr. Hj. Sri Hariani, M.Si., selaku dekan Fakutas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Ir. Fachrul Kurniawan, M.MT., IPU, selaku Ketua Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Dr. Muhammad Faisal, M.T selaku Dosen Pembimbing I yang telah dengan sabar memberikan arahan baik dalam penulisan hingga program yang dibuat dalam menyelesaikan skripsi ini.
5. Dr. Irwan Budi Santoso, M.Kom. selaku Dosen Pembimbing II yang telah memberikan bimbingan, arahan serta bantuan dalam terwujudnya karya tulis skripsi ini.
6. Hani Nurhayati, M.T selaku ketua penguji, Tri Mukti Lestari, M.Kom selaku penguji I yang telah meluangkan waktunya untuk menguji dan dengan sabar memberi arahan serta saran dalam menyelesaikan skripsi ini.

7. Dr. Ir. Mokhammad Amin Hariyadi, M.T selaku Dosen Wali yang telah memberikan arahan dalam proses perkuliahan.
8. Khadijah Fahmi Hayati Holle, M.Kom dan Tri Mukti Lestari, M.Kom yang telah mengajarkan *Natural Language Processing* dan *Information Retreival* sehingga muncul ide untuk melakukan penelitian di bidang tersebut.
9. Nia Faricha, S.Si selaku Admin Program Studi Teknik Informatika yang dengan sabar membantu, memberikan arahan informasi terkait perkuliahan.
10. Segenap Dosen, Laboran dan jajaran pada Program Studi Teknik Informatika yang telah memberikan bimbingan dan bantuan selama studi.
11. Kedua orangtua penulis Bapak Sumarji dan Ibu Marni yang selalu memberi dukungan dan doa serta selalu memberikan yang terbaik untuk kelancaran putrinya dalam pendidikan.
12. Kakak laki-laki penulis Muhammad Aji Cahyono dan Keponakan penulis Abimanyu yang selalu memberi semangat dan menghibur di tengah kegiatan kuliah yang padat.
13. Sahabat penulis Eka Mira Novita Subroto, Shafira Halmahera, Radina Mutia Haira, Aninda Rizky Hartanti, Suci Wulandari, dan Popi Merkuri yang selalu memberi semangat kepada penulis.
14. Seluruh pihak yang telah terlibat secara langsung maupun tidak langsung dalam proses penyusunan skripsi sejauh ini.

Akhir kata, penulis mengakui bahwa penulisan pada skripsi ini masih banyak kekurangan. Penulis berharap semoga skripsi ini diterima sebagai amal ibadah yang tulus dan bermanfaat di sisi Allah Subhanahu Wa Ta'ala. Semoga

karya ini menjadi bagian dari kontribusi yang tak terputus dalam rangka memperkuat dan mengembangkan ilmu pengetahuan, serta melaksanakan tugas sebagai hamba Allah yang berkomitmen.

Wassalamualaikum Warahmatullahi Wabarakatuh.

Malang, 2 April 2025

Penulis

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

ABSTRAK

Cahyani, Vivin Octavia. 2025. **Peringkasan Teks Ekstraktif Karya Ilmiah Mahasiswa Menggunakan *Fuzzy C-Means* dan *Vector Space Model*.** Skripsi. Program Studi Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. Muhammad Faisal, M.T (II) Dr. Irwan Budi Santoso, M.Kom.

.

Kata Kunci: Fuzzy C-Means, Vector Space Model, Peringkasan Teks.

Artikel jurnal ilmiah terus meningkat setiap tahunnya, sering kali mempersulit pembaca dalam menyaring informasi inti secara efisien. Informasi yang kurang efisien membuat pembaca harus membaca ulang artikel sehingga memerlukan banyak waktu. Oleh karena itu, dibutuhkan sebuah alat untuk menemukan inti informasi dari artikel jurnal ilmiah secara cepat dan efisien. Untuk mengatasi masalah ini, peringkasan teks otomatis diperlukan, khususnya untuk artikel jurnal ilmiah berbahasa indonesia. Penelitian ini mengembangkan sistem peringkasan teks otomatis menggunakan metode *Fuzzy C-Means* dan *Vector Space Model* menggunakan pembobotan kalimat TF-IDF (*Term Frequency Invers Document Frequency*). Evaluasi sistem menggunakan metrik ROUGE-1 dan ROUGE-2. Hasil pengujian menunjukkan bahwa sistem terbaik, pada tingkat kompresi 30% serta menggunakan *stemming* memberikan hasil terbaik dan seimbang, dengan rata-rata ROUGE-1 *Precision* 0.5331, *Recall* 0.5034, *F1-Score* 0.4975 dan *Accuracy* 0.5183. Hasil penelitian ini menunjukkan bahwa model dengan *stemming* lebih disarankan untuk menghasilkan ringkasan yang lebih relevan dan akurat pada tingkat kompresi yang lebih tinggi.

ABSTRACT

Cahyani, Vivin Octavia. 2025. **Extractive Text Summarization of Student Scientific Works Using Fuzzy C-Means and Vector Space Model.** Undergraduate Thesis. Informatics Engineering Study Program, Faculty of Science and Technology Maulana Malik Ibrahim State Islamic University Malang. Supervisor: (I) Dr. Muhammad Faisal, M.T (II) Dr. Irwan Budi Santoso, M.Kom.

Keywords: Fuzzy C-Means, Vector Space Model, Text Summarization.

Scientific journal articles increase yearly, often making it difficult for readers to sift through the core information efficiently. Inefficient information makes the reader have to re-read the article, which takes a lot of time. Therefore, a tool is needed to find the core information of scientific journal articles quickly and efficiently. To solve this problem, automatic text summarization is needed, especially for Indonesian scientific journal articles. This research develops an automatic text summarization system using the Fuzzy C-Means method and Vector Space Model using TF-IDF (Term Frequency Inverse Document Frequency) sentence weighting. The system evaluation uses ROUGE-1 and ROUGE-2 metrics. The test results show that the best system, at 30% compression level and using stemming gives the best and balanced results, with an average ROUGE-1 Precision of 0.5331, Recall of 0.5034, F1-Score 0.4975, and Accuracy of 0.5183. The results of this study show that the model with stemming is recommended to produce more relevant and accurate summaries at higher compression levels.

مستخلص البحث

جاهياني، فيفين أوكتافيا. 2025. تلخيص النصوص الاستخراجية للأوراق العلمية للطلاب الجامعي باستخدام خوارزمية العنقدة الضبابية والوسطاء المتعددين ونموذج فضاء المتجه. البحث الجامعي. قسم الهندسة المعلوماتية، كلية العلوم والتكنولوجيا .بجامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرف الأول: د. محمد فيصل، الماجستير. المشرف الثاني: د .إيروان بودي سانتوسو، الماجستير

.**الكلمات الرئيسية:** خوارزمية عنقدة ضبابية ووسطاء متعددين، نموذج فضاء متجه، تلخيص نص

تستمر مقالات المجلة العلمية في الزيادة كل عام، مما يجعل من الصعب على القراء في التدقيق بكفاءة في المعلومات الأساسية. المعلومات الأقل كفاءة تجعل القراء يضطرون إلى إعادة قراءة المقالة، لذا يستغرق الأمر الكثير من الوقت. لذلك، هناك حاجة إلى أداة للعثور بسرعة على المعلومات الأساسية من مقالات المجلة العلمية. لحل هذه المشكلة، هناك حاجة إلى تلخيص تلقائي للنص، خاصة بالنسبة لمقالات المجلات العلمية الإندونيسية. طور هذا البحث نظاما آليا لتلخيص النص باستخدام طريقة خوارزمية العنقدة الضبابية والوسطاء المتعددين ونموذج فضاء المتجه من خلال ترجيح كلمة TF-IDF (معدل تكرار الكلمة الواحدة في المستند الواحد). استخدم تقييم النظام مقاييس ROUGE-1 وROUGE-2 . أظهرت نتائج الاختبار أن النظام بنسبة ضغط 30٪ وباستخدام جذع الكلمة أعطى أفضل النتائج وأكثرها توازنا، بمتوسط لـ ROUGE-1 هو الثبات 0.53331، والاستدعاء 0.5034، ودرجة ف1 0.4975، ودقة 0.5183. أشارت نتائج هذا البحث إلى أن النماذج ذات جذع الكلمة موصى بها أكثر لإنتاج ملخصات أكثر صلة ودقة بمستويات ضغط أعلى.

.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Artikel jurnal ilmiah adalah sebuah karya tulis yang terstruktur menggunakan kaidah-kaidah keilmuan, berisi tentang permasalahan faktual atau nonfiksi yang dimuat dalam jurnal, majalah, atau berita dengan tujuan untuk menyampaikan sebuah fakta dan menawarkan solusi dari suatu permasalahan (Salma, 2022). Saat ini, artikel jurnal ilmiah dapat ditemukan secara online. Banyak artikel yang diterbitkan di berbagai situs web, dan dapat dibaca kapan saja serta dimana saja.

Perkembangan teknologi informasi yang begitu pesat membuat peningkatan dokumen teks digital secara signifikan. Ilmu pengetahuan yang semakin maju membuat banyak penelitian yang dikembangkan, salah satu buktinya dengan banyaknya publikasi artikel jurnal ilmiah, yang mengharuskan pembaca untuk membaca keseluruhan isi dari artikel. Publikasi artikel jurnal ilmiah memudahkan mahasiswa untuk mendapatkan referensi penelitian, dan memudahkan dosen untuk proses penilaian artikel jurnal ilmiah lewat hasil publikasi mahasiswa. Publikasi yang dilakukan oleh dosen, mahasiswa, maupun peneliti meningkat setiap tahun. Pada tahun 2022, China berhasil menduduki peringkat pertama penghasil Artikel Jurnal Ilmiah Terindeks Scopus terbanyak, dengan jumlah total 1.004.745 jurnal (Wahyono, 2023). Banyaknya artikel jurnal ilmiah yang dipublikasikan mencakup berbagai topik, dengan publikasi yang terus meningkat setiap tahunnya.

Seiring dengan meningkatnya jumlah publikasi artikel jurnal ilmiah setiap tahunya, pembaca termasuk mahasiswa, peneliti, dan akademisi, sering kali menghadapi kesulitan dalam menyaring informasi yang relevan dari banyaknya artikel yang diterbitkan. Pencarian informasi inti dari setiap artikel menjadi tugas yang semakin kompleks, terutama ketika artikel-artikel tersebut memuat informasi yang mendalam dan spesifik, sehingga membutuhkan waktu yang tidak sedikit untuk memahami secara keseluruhan. Ketidakefektifan ini menghambat kemampuan pembaca untuk mengakses dan memanfaatkan literatur ilmiah secara efektif. Kebutuhan untuk meringkas artikel jurnal ilmiah muncul, untuk mengatasi masalah *overload* informasi bagi mahasiswa yang sedang menyusun karya ilmiah, seperti skripsi atau tesis, yang membutuhkan akses dengan cepat untuk mendapatkan informasi inti dari berbagai artikel jurnal ilmiah. Namun, meskipun tingginya minat baca dan kebutuhan akan referensi dari artikel jurnal ilmiah, pembaca sering kali kewalahan oleh banyaknya informasi yang tersedia. Oleh karena itu, pengembangan sistem peringkasan teks otomatis yang efektif menjadi sangat penting, untuk membantu pembaca dalam memperoleh informasi inti secara cepat dan efisien tanpa harus membaca keseluruhan isi artikel.

Peringkasan teks otomatis merupakan sebuah cara peringkasan teks dengan menggunakan bantuan perangkat lunak sehingga menghasilkan informasi inti dari sebuah teks panjang yang cepat dan efisien (Dewi & Widiastuti, 2022). Tujuannya adalah memudahkan pembaca dalam memahami inti isi artikel tanpa perlu membaca keseluruhan teks.

Jenis peringkasan teks otomatis berdasarkan jumlah dokumen yang digunakan dibagi menjadi dua dengan peringkasan teks *single document* dan peringkasan teks *multi document.* Sedangkan berdasarkan hasil *output*nya dibagi juga menjadi dua yaitu, peringkasan teks ekstraktif (*extractive summarization*) dan peringkasan teks abstraktif (*abstractive summarization*). *Extractive summarization* merupakan metode meringkas dengan mengambil kalimat-kalimat yang dianggap penting dan menjadikannya satu kesatuan sehingga mendapatkan hasil ringkasan, sedangkan ringkasan teks abstraktif (*abstractive summarization*) dengan membuat kata-kata baru yang dapat merepresentasikan hasil ringkasan dalam kalimat atau bisa dikatakan dengan cara parafrase (Halimah et al., 2022).

Dalam bidang ilmu agama islam kita senantiasa diwajibkan untuk menuntut ilmu dengan baik dan menyebarkan ilmu tersebut supaya bermanfaat bagi orang disekitar kita sebagaimana diterangkan dalam firman Allah SWT pada surat Al Ankabut ayat 43:

وَتِلْكَ الأَمْثَالُ نَضْرِبُهَا لِلنَّاسِ وَمَا يَعْقِلُهَا إِلا الْعَالِمُونَ

“Dan perumpamaan-perumpamaan ini Kami buatkan untuk menusia; dan tiada yang memahaminya kecuali orang-orang yang berilmu.” (QS. al ankabut : 43)

Penjelasan menurut tafsir jalalayn mengenai ayat ini yaitu (Dan perumpamaan-perumpamaan ini) yang ada didalam al-quran (Kami buatkan) Kami jadikan (untuk manusia dan tiada yang memahaminya) yang mengerti akan perumpamaan-perumpamaan ini (kecuali orang-orang yang berilmu) yakni, orang-orang yang mau berfikir. Dalam hubungannya dengan manusia, informasi yang terkandung dalam suatu dokumen artikel jurnal ilmiah yang berisi sebuah ilmu bagi orang yang mau

berfikir dan mengambil manfaat dari ilmu tersebut maka pahala bagi orang yang menulis karena melalui tulisanya dapat membantu orang yang membaca dan memahami isinya. Karena dengan akal, manusia bisa berfikir melalui ilmu. Sehingga sistem peringkasan teks otomatis artikel jurnal ilmiah diharapkan mampu menjadi salah satu alat untuk memudahkan masyarakat dalam mendapatkan inti isi informasi dari sebuah artikel jurnal ilmiah.

Sistem peringkasan teks dapat dilakukan secara otomatis dengan memanfaatkan kecerdasan buatan (*Artifilicial Intelligent*) yang dapat melakukan peringkasan teks artikel jurnal ilmiah berdasarkan model yang sudah diberikan. Salah satu model yang diberikan adalah *Vector Space Model* dan *Fuzzy C-Means* dimana *Vector Space Model* digunakan penelitian oleh (Utomo et al., 2022) sebagai pembobotan nilai pada setiap kata yang ada di artikel, penelitian ini menghasilkan sebuah paragraf ringkasan yang diambil dari beberapa kalimat yang memiliki nilai kemiripan dengan judul yang paling tinggi dari 104 kalimat terdapat 5 kalimat yang paling tinggi nilainya, dari 5 kalimat tersebut dijadikan satu paragraf untuk hasil peringkasannya.

Fuzzy C-Means merupakan sebuah algoritma yang dikembangkan oleh Dunn pada tahun 1973 dan pada tahun 1981 diperbaiki oleh Bezdek yang banyak digunakan untuk pengenalan sebuah pola (Bharathi et al., 2016). *Fuzzy C-Means* dapat menjadikan setiap objek dari beberapa *clusters* menjadi pilihan yang tepat, mampu mengelompokkan data yang luas, kokoh terhadap data outlier serta sederhana dan mudah diterapkan (V. K. Singh et al., 2011). Implementasi *Fuzzy C-Means* untuk meringkas dokumen jurnal berbahasa inggris dengan input teks

berformat pdf. dan menggunakan pembobotan kalimat menggunakan metode TF-IDF berhasil mendapatkan nilai rata-rata *recall, precision,* dan *f-measure* masing-masing sebesar 47,8%, 32% dan 37,04%. Pada tahap evaluasi, nilai *recall* tertinggi sebesar 65%, sedangkan untuk *presicion* dan *f-measure* masing-masing sebesar 30% dan 41,05% untuk penelitian yang pernah dilakukan oleh (Irfan et al., 2017).

Pada keberhasilan dari penelitian yang menggunakan kedua metode tersebut dalam meringkas teks, dokumen yang digunakan masih banyak menggunakan artikel berita berbahasa indonesia, artikel berita berbahasa inggris, dan jurnal berbahasa inggris. Sedangkan untuk *dataset* seperti dokumen artikel jurnal ilmiah karya mahasiswa berbahasa indonesia belum banyak dilakukan peringkasan. Pada penelitian yang dilakukan oleh (Suwija Putra et al., 2021) hasil evaluasi menunjukkan bahwa *Fuzzy C-Means* dengan pembobotan kalimat TF-ISF belum menghasilkan kinerja optimal karena pembobotan kalimat menggunakan TF-ISF tidak cukup representatif untuk menentukan relevansi kalimat dalam dokumen yang sangat kompleks seperti essai, dengan celah tersebut penelitian ini akan mengisinya menggunakan metode *Fuzzy C-Means* dengan *Vector Space Model* yang mengukur kemiripan semantik antar kalimat sehingga dapat melakukan pengelompokkan kalimat yang lebih akurat berdasarkan kemiripan antar teks. Berdasarkan alasan tersebut diangkatlah penelitian ini tentang peringkasan teks otomatis untuk artikel jurnal ilmiah karya mahasiswa berbahasa indonesia, sistem yang akan dibuat nantinya perlu evaluasi untuk mengetahui seberapa baik performanya dalam meringkas. Pengukuran evaluasi yang digunakan akan menggunakan ROUGE dimana akan diukur nilai *Precision, Recall, F-Measure* pada metode *Vector Space*

Model sebagai pembobotan kalimat serta metode *Fuzzy C-Means* sebagai salah satu pendekatan dalam meringkas teks.

1.2 Rumusan Masalah

Seberapa baik performa dari hasil sistem peringkasan teks artikel jurnal ilmiah karya mahasiswa menggunakan metode *Fuzzy C-Means* dan *Vector Space Model* ?

1.3 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Penelitian akan dilakukan dengan pendekatan ekstraktif dan inputan *single* dokumen.
2. *Dataset* yang digunakan dalam penelitian ini hasil dari pengumpulan artikel jurnal ilmiah yang telah dipublikasikan oleh mahasiswa pada *Repository* UIN Malang, *Google Schoolar*, dan *Sciene Tecnology Index* (SINTA).

1.4 Tujuan Penelitian

Mengukur seberapa baik performa sistem peringkasan teks otomatis artikel jurnal ilmiah karya mahasiswa menggunakan metode *Fuzzy C-Means* dan *Vector Space Model* yang dievaluasi menggunakan ROUGE untuk mendapatkan nilai *precision, recall, f-measure* dan *accuracy.*

1.5 Manfaat Penelitian

Sistem peringkasan teks dapat dimanfaatkan oleh akademisi dan masyarakat umum yang membutuhkan peringkasan teks. Sistem akan membatu pengguna untuk menemukan inti informasi secara cepat, tanpa harus membaca keseluruhan isi dari artikel. Sistem juga membantu untuk menghemat waktu pembaca dengan memberikan ringkasan singkat dari teks artikel yang panjang.

BAB II

STUDI PUSTAKA

2.1 Penelitian Terkait

Penelitian dahulu pernah dilakukan oleh (Gunawan et al., 2023) dengan judul "*Maximum Marginal Relevance and Vector Space Model for Summarizing Students' Final Project Abstracts*". Pada penelitian tersebut menggunakan *dataset* abstrak yang diambil dari 200 tugas akhir mahasiswa dan dokumen skripsi yang diujicobakan pada sistem peringkasan teks. Tahapan untuk proses peringkasan tersebut yaitu *preprocessing* meliputi *sentence breaking, case folding, tokenizing, filtering,* dan *stemming,* menghitung bobot kalimat menggunakan TF-IDF yang dihitung menggunakan relevansi kueri *Vector Space Model,* menghitung kesamaan antar kalimat menggunakan *cosine similarity* dan yang terakhir menggunakan *Maximum marginal relevance* (MRR) untuk ekstraksi ringkasan. Hasil percobaan membandingkan ringkasan dari sistem dan ringkasan referensi yang dibuat oleh pakar menghasilkan skor rata-rata *precision* 88%, *recall* 61%, dan *f-measure* 70%. Perbedaan dengan penelitian ini terletak pada *dataset* yang digunakan pada penelitian terkait adalah abstrak dari tugas akhir sedangkan penelitian ini menggunakan artikel jurnal ilmiah. Penelitian terkait menggunakan MMR untuk derajat kemiripan antar bagian teks dari hasil relevansi kueri sedangkan pada penelitian ini menggunakan *fuzzy c-means* untuk mengelompokkan kalimat berdasarkan kemiripannya.

Penelitian dahulu pernah dilakukan oleh (R. Singh & Singh, 2021) dengan judul "*Text Similarity Measure in News Articles by Vector Space Model Using NLP*". Pada penelitian tersebut menggunakan artikel berita yang didapatkan dari *Google News* berbahasa india yang diterjemahkan ke bahasa inggris dan akan dibandingkan dengan artikel berita berbahasa inggris. Sistem yang digunakan pada penelitian terkait berfokus pada perbandingan tiga metode yang berbeda untuk memperkirakan kesamaan semantik di antara dua artikel berita yang hampir sama topiknya tetapi berbeda pada bahasa (Hindi dan Inggris), ketiga metode tersebut adalah *Cosine similarity with tf-idf vectors, similarity of Jaccard with tf-idf vectors,* dan *Bag of word Euclidean distance.* Ketiga metode yang digunakan menunjukkan hasil akurasi lebih besar pada metode *cosine similarity with tf-idf vectors* dengan perolehan nilai rata-rata *accuracy, recall,* dan *f-measure* masing-masing sebesar 81,25%, 100% dan 76,92%. Penelitian rujukan menggunakan tiga perbandingan metode vektor kemiripan tanpa meringkas sedangkan pada penelitian ini menggunakan *Vector Space Model* yang akan digunakan untuk meringkas teks. *Dataset* yang digunakan pada penelitian rujukan menggunakan data artikel berita dari *Google News* menggunakan dua bahasa (Hindi dan Inggris) sedangkan pada penelitian ini menggunakan artikel jurnal ilmiah.

Penelitian dahulu pernah dilakukan oleh (Setiawan et al., 2022) dengan judul "*Sentiment Summmarization* Evaluasi Pembelajaran Menggunakan Algoritma *LSTM (Long Short Term Memory)*" pada penelitian ini menggunakan *dataset* yang berasal dari hasil evaluasi pembelajaran dosen diisi oleh mahasiswa yang digunakan untuk bahan refleksi diri dosen supaya meningkatkan kinerja dalam

pembelajaran kedepannya. Dari evaluasi itu diperlukan teknik analisis untuk mengklasifikasi dari teks panjang menjadi teks padat yang informatif. Pada penelitian tersebut diterapkan teknik abstraktif dengan algoritma *Long Short Term Memory (LSTM)* digunakan untuk proses klasifikasi sentimen dan peringkasan teks. Berdasarkan hasil pengujian *confusion matrix*, sistem memperoleh nilai akurasi sebesar 0,902 dan nilai *f-measure* sebesar 0,921. Pada pengujian menggunakan ROGUE diperoleh hasil evaluasi positif 0,16 dan hasil evaluasi negatif sebesar 0,2. Penelitian rujukan menggunakan data dari hasil evaluasi dosen yang diisi oleh mahasiswa, sedangkan penelitian ini menggunakan data artikel jurnal ilmiah. Pada penelitian sebelumnya, algoritma yang digunakan *Long Short Term Memory* sementara penelitian ini menerapkan metode *Vector Space Model* dan *Fuzzy C-Means.*

Penelitian dahulu pernah dilakukan oleh (Samosir et al., 2022) dengan judul "BESKlus : BERT *Extractive Summarization with K-Means Clustering in Scientific Paper*". Penelitian tersebut menggunakan model BESKlus, sebuah model sistem peringkasan teks otomatis ekstraktif dengan *contextual embedding* menggunakan kombinasi *Sentence*-BERT dan *K-Means Clustering.* Hasil dari evaluasi menggunakan ROUGE yaitu ROUGE-L dan BERTScore, untuk skor pada ROUGE-L nilai rata-rata *recall, precision* dan *F1* masing-masing sebagai berikut 17%, 18%, dan 18% sedangkan untuk skor pada BERTScore nilai rata-rata *recall, precision,* dan *F1* masing-masing sebagai berikut 88,35%, 86,10%, dan 87,2%. Penelitian rujukan menggunakan metode peringkasan teks BERT dan *K-Means Clustering* sedangkan pada penelitian ini menggunakan *Vector Space Model* dan

Fuzzy C-Means. Data untuk uji coba pada penelitian rujukan berasal dari Kaggle kumpulan makalah jurnal ilmiah NeurIPS tahun 1987-2019 berbahasa inggris sedangkan pada penelitian ini menggunakan data artikel jurnal ilmiah yang dipublikasikan oleh mahasiswa berbahasa indonesia.

Penlitian yang pernah dilakukan oleh (Halimah et al., 2022) yang berjudul "Peringkasan Teks Otomatis *(Automated text summarization)* pada artikel berbahasa indonesia menggunakan algoritma *Lexrank*". Penelitian tersebut melalui beberapa tahapan, dimulai dari proses *preprocessing.* Setelah itu, dilakukan perhitungan bobot kalimat menggunakan metode TF-IDF. Data yang digunakan untuk penelitian tersebut berasal dari 300 artikel yang diambil dari internet. Hasil penelitian menunjukkan bahwa tingkat kompresi 50%, nilai *f-measure* rata-rata pada metriks ROUGE-1, ROUGE-2 dan ROUGE-L berturut-turut adalah 67,53%, 59,10% dan 67,05%. Sedangkan pada tingkat kompresi 30% nilai *f-measure* rata-rata untuk metriks ROUGE-1, ROUGE-2 dan ROUGE-L berturut-turut sebesar 55,82%, 45,51%, dan 54,76%. Perbedaan yang tampak dari penelitian terletak pada metode peringkasan yang digunakan, penelitian rujukan menggunakan metode *Lexrank,* sedangkan pada penelitian ini menggunakan metode *Vector Space Model* dan *Fuzzy C-Means.* Data yang digunakan pada penelitian rujukan menggunakan 300 artikel dari internet, sedangkan pada penelitian ini menggunakan kumpulan artikel jurnal ilmiah.

Penelitian dahulu pernah dilakukan oleh (Suwija Putra et al., 2021) dengan berjudul "*Extractive Text Summarization of Student Essay Assignment Using Sentence Weigh Features and Fuzzy C-Means*". Penelitian tersebut membuat

aplikasi peringkasan teks dokumen dari tugas essai mahasiswa yang memudahkan dosen dalam memberikan penilaian. Peringkasan dokumen teks tugas esai secara otomatis menggunakan metode *Fuzzy C-Means* dengan fitur bobot kalimat dengan TF-ISF, sistem berhasil meringkas teks dengan rata-rata nilai evaluasi *presicion, recall, accuracy,* dan *f-measure* yang didapat masing-masing sebesar 0,52, 0,54, 0,70 dan 0,52. Perbedaan penelitian terletak pada metode peringkasan teks yang digunakan walaupun pada penelitian rujukan juga menggunakan *fuzzy c-means* tetapi tidak menggunakan *vector space model,* sedangkan pada penelitian ini menggunakan *vector space model* dan *fuzzy c-means. Dataset* yang digunakan pada penelitian rujukan menggunakan artikel tugas essai mahasiswa, sedangkan data uji pada penelitian ini menggunakan artikel jurnal ilmiah yang didapat dari berbagai sumber artikel jurnal di internet.

Penelitian dahulu dilakukan oleh (Aditya & Wiyono, 2023) yang berjudul "Pengembangan Fitur Peringkasan Artikel Otomatis pada Media Online Satukanal". Penelitian ini membuat sebuah *website* peringkasan teks berita online untuk membantu mengatasi masalah waktu dan biaya dalam pembuatan ringkasan berita secara manual. Pengembangannya menggunakan metode *Fuzzy C-Means* (FCM) untuk *clustering* dan *Technique for Order Preference by Similarity to Ideal Solution* (TOPSIS). Program ini juga menggunakan metode *Rapid Application Development* (RAD), yang melibatkan pengguna dalam proses pengembangan pelatihan untuk memastikan sistem dapat digunakan secara maksimal. Program ini berhasil meningkatkan produktivitas berita di media online Satukanal dan memberikan solusi terhadap masalah pengelolaan berita dalam jumlah besar.

Penjelasan mengenai penelitian sebelumnya yang menggunakan metode diantaranya *Maximum marginal Relevance (MMR), Long Short Term Memory (LSTM), Lexrank, K-Means Clustering, Vector Space Model (VSM),* dan BERT untuk peringkasan teks. Terdapat juga penelitian yang menggunakan *Fuzzy C-Means* untuk menunjang penelitian peringkasan teks yang berhasil menggabungkannya dengan fitur bobot kalimat dan beberapa penelitian tersebut banyak yang menggunakan objek penelitiannya menggunakan artikel hasil tugas esai. Keberhasilan penelitian yang menggunakan metode *Fuzzy C-Means* membuat peneliti tertarik untuk menggabungkannya dengan metode *Vector Space Model* yang dapat menghasilkan peringkasan teks lebih baik. Hasil dari peringkasan nantinya akan di evaluasi menggunakan matrik *Recall Oriented Underresearch for Gisting Evaluation* (ROUGE) untuk mengetahui seberapa baik hasil sistem peringkasan yang dibuat. Terdapat dua pembaharuan pada penelitian ini, pertama penggunaan *dataset* dari hasil pengumpulan artikel jurnal ilmiah yang tersedia dan kedua menggunakan metode *Vector Space Model* untuk membuat vektor kalimat yang dikombinasikan dengan metode *Fuzzy C-Means* digunakan pada pengelompokan kalimat membentuk *cluster*.

2.2 Artikel Jurnal Ilmiah

Artikel jurnal ilmiah merupakan karya ilmiah yang biasa di publikasikan oleh mahasiswa, dosen dan peneliti yang digunakan untuk sumber rujukan penelitian. Menurut (Jatmiko, 2015) artikel ilmiah adalah sebuah tulisan yang berisi ide, gagasan dan hasil pemikiran seseorang atau sekelompok orang yang telah dihasilkan melalui proses penelitian, pengamatan dan evaluasi. Artikel jurnal ilmiah

disusun dalam bentuk laporan tertulis sesuai dengan metodologi, sistem standar yang telah disepakati sehingga dapat dipertanggungjawabkan secara ilmiah dapat diuji kebenarannya. Artikel ini dapat dipublikasikan pada jurnal ilmiah nasional atau internasional. Artikel jurnal ilmiah dapat diakses dengan mudah di situs *web*, semakin berkembangnya teknologi banyak penelitian yang mempublikasikan hasil penelitianya ke publik supaya dapat memberi informasi terbaru tentang perkembangan penelitian saat ini dan dapat mengembangkan ilmu pengetahuan kedepannya.

2.3 *Vector Space Model*

Vector Space Model adalah suatu metode yang dimanfaatkan untuk mengetahui seberapa dekat atau mirip suatu *term* dengan menggunakan teknik pemberian bobot pada kata *(term).* Dalam konteks ini, sebuah dokumen dapat dianggap sebagai vektor yang memiliki panjang *(magnitude)* dan arah *(direction).* Dalam model ini, setiap kata diwakili sebagai salah satu dimensi dalam ruang vektor. Kompatibilitas vektor dokumen dan vector *query* menentukan relevansi dokumen dengan *query* (Baeza-yates & Ribeiro-neto, n.d.). Sebuah kerangka parsial adalah mungkin diberikan oleh *Vector Space Model.* Dalam merepresentasikan sebuah *vector* diperlukan bobot *term* dari dokumen maupun *query*.. *Term* bisa berupa kata, frase, atau unit hasil *indexing* dalam sebuah dokumen sebagai gambaran isi dari dokumen tersebut.

Rumus dokumen dan *query* direpresentasikan sebagai vektor.

$$d_{j\,=(w_{1,j},w_{2,j,}\ldots\ldots,w_{t,j})} \qquad (2.1)$$

$$q = (w_{1,q},\ w_{2,q}, \dots\dots, w_{n,q}) \qquad (2.2)$$

Sebuah dokumen dj dan sebuah *query* q direpresentasikan sebagai vektor t-dimensi seperti pada Gambar 2.1.

Gambar 2. 1 Kosinus diadopsi dari θ sebagai sim dj, q

Dalam *Vector Space Model* kumpulan dokumen direpresentasikan sebagai sebuah matrik *term frequency.* Setiap sel dalam matrik sesuai dengan bobot yang diberikan pada *term* dokumen. Jika bernilai nol maka *term* tersebut tidak ada di dalam dokumen. Gambar 2.2 menunjukkan matrik *term frequency* dengan n dokumen dan t *term.*

	T_1	T_2	T_3	$T_{...}$	T_t
D_1	W_{11}	W_{21}	W_{31}	$\square_{...}$	T_{t1}
D_2	W_{12}	W_{22}	W_{32}	$\square_{...}$	T_{t2}
D_3	W_{13}	W_{23}	W_{33}	$\square_{...}$	T_{t3}
$D_{...}$	$\square_{...}$	$\square_{...}$	$\square_{...}$	$\square_{...}$	$\square_{\square}$
D_n	W_{1n}	W_{2n}	W_{3n}	$\square_{...}$	T_{tn}

Gambar 2. 2 Metrik term frequency

Setiap dimensi sesuai dengan istilah yang terpisah. Jika jangka terjadi dalam dokumen, nilai didalam vektor adalah non-biner. Ada beberapa cara yang beda dari komputasi nilai, juga dikenal dengan (istilah) berat atau bobot, telah dikembangkan. Metode pembobotan yang banyak digunakan secara umum dalam *Vector Space*

Model yaitu *Term Frequency Inverse Document Frequency* (TF-IDF). Proses perhitungan tahap *term frequency* menggunakan persamaan (2.3)

$$tf = tf_{ij} \quad (2.3)$$

Tf adalah *term frequency* dan $tf_{i,j}$ adalah banyaknya kemunculan kata *(term)* t_i dalam dokumen d_j. Nilai *term frequency* dihitung berdasarkan banyaknya kemunculan *term* t_i dalam sebuah dokumen d_j.

Untuk perhitungan *Invers Document Frequency* menggunakan persamaan (2.4)

$$idf_i = log \frac{N}{df_i} \quad (2.4)$$

IDF (*invers document frequency*) didefinisikan sebagai nilai yang menunjukkan seberapa banyak dokumen dalam koleksi yang memuat *term* t_i dengan *N* sebagai jumlah total dokumen yang diambil oleh sistem, dan *dfi* sebagai jumlah dokumen tempat *term ti* muncul. Perhitungan idf_i bertujuan untuk mengidentifikasi frekuansi kemunculan *term* yang dicari (df_i) dalam dokumen lain didalam basis data (korpus).

Nilai *term frequency invers document frequency (tfidf),* dihitung menggunakan persamaan (2.5)

$$W_{ij} = tf_i . log (\frac{N}{df_i} \quad (2.5)$$

Dengan W_{ij} menunjukkan bobot dokumen, *N* menunjukkan jumlah dokumen yang diambil oleh sistem, $tf_{i,j}$ menunjukkan banyaknya kemunculan kata *(term)* pada dokumen (d_j) dan (df_i) menunjukkan banyaknya dokumen dalam

koleksi yang mengandung *term* t_i. Untuk menghitung bobot dokumen *(W_{ij})* harus menghitung hasil perkalian atau kombinasi dari frekuensi *term* dokumen *($tf_{i,j}$)* dan frekuensi *invers* dokumen *(df_i).*

Perhitungan dari Jarak *query* menggunakan persamaan (2.6) dan dokumen, menggunakan persamaan (2.7)

$$|q| = \sqrt{\sum_{j=1}^{t} \left(W_{i,q}\right)^2} \tag{2.6}$$

Nilai mutlak q atau *($|q|$)* mewakili jarak *query*, sementara $W_{i,q}$ menggambarkan bobot dari *query* pada dokumen ke-i. Nilai *($|q|$)* diperoleh untuk menentukan jarak *query* terhadap bobot *query* dokumen ($W_{i,q}$) yang sudah diekstraksi oleh sistem. Perhitungan jarak *query* ini dilakukan menggunakan rumus akar dari jumlah kuadrat *query.*

$$\left|d_j\right| = \sqrt{\sum_{i=1}^{t} \left(W_{i,j}\right)^2} \tag{2.7}$$

Apabila *dj* merupakan jarak dokumen, dan $W_{i,,j}$ menggambarkan bobot dari dokumen ke-i, maka jarak dokumen (|d|) dapat diperoleh untuk menngetahui bobot dokumen *($W_{i,j}$)* yang diekstraksi sistem. Proses perhitungan jarak dokumen ini dilakukan dengan menggunakan rumus akar jumlah kuadrat dari dokumen.

Perhitungan pengukuran Similaritas *query document (inner product),* dengan persamaan (2.8)

$$sim\left(q, d_j\right) = \sum_{i=1}^{t} W_{iq} \ . W_{ij} \tag{2.8}$$

Bobot *term* dalam dokumen dilambangkan dengan W_{ij} sedangkan W_{iq} adalah bobot *query*, dan *sim(q, d_j)* merepresentasikan hubungan antara *query* dengan dokumen. Bobot hubungan antara *query* dan dokumen atau *inner product (sim (q, d_j))* digunakan untuk menentukan bobot, baik melalui perkalian bobot *q* dengan bobot dokumen maupun dengan memanfaatkan bobot *term* dokumen *(W_{ij})* dan bobot *query (W_{iq}).*

Pengukuran untuk menghitung nilai kosinus sudut antara dua vektor atau *Cosine Similarity* menggunakan persamaan (2.9)

$$Sim(q,\ d_j) = \frac{q.d_j}{|q| * |d_j|} = \frac{\sum_{i=1}^{t} W_{iq} . W_{ij}}{\sqrt{\sum_{j=1}^{t} (W_{iq})^2} \quad * \quad \sqrt{\sum_{i=1}^{t} (W_{ij})^2}} \qquad (2.9)$$

Tingkat kemiripan antara pernyataan dan dokumen, yang dikenal sebagai *Sim (q, d_j)*, sebanding dengan hasil perkalian antara bobot pernyataan *(q)* dan bobot dokumen *(d_j)* serta berbanding terbalik dengan hasil akar dari jumlah kuadrat *(|q|)* dan jumlah kuadrat dokumen *(d_j)*. Proses perhitungan similaritas akan menghasilkan bobot dokumen yang lebih besar dibandingkan nilai dari perhitungan *inner product* atau yang mendekati angka 1. Setelah bobot (*W*) pada tiap dokumen diketahui, maka dilakukan proses penyortiran. Semakin tinggi nilai *W* semakin besar tingkat persamaan antara dokumen dengan kata kunci yang dicari. Hal yang sama juga berlaku sebaliknya.

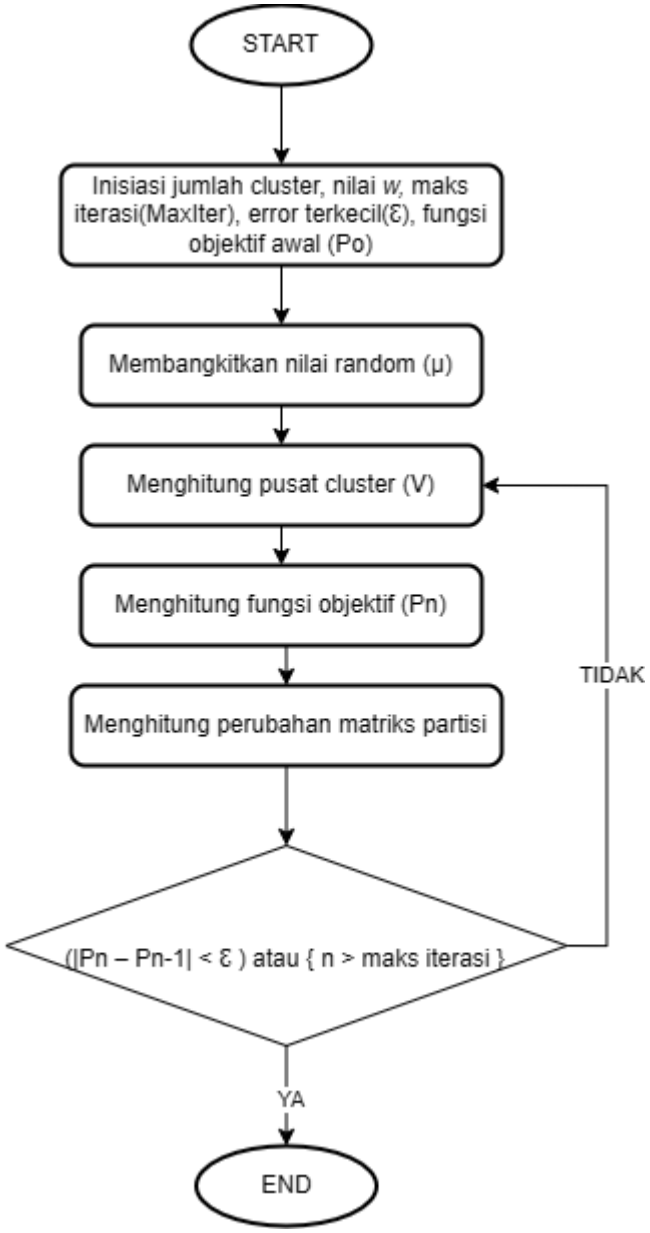
2.4 *Fuzzy C-Means*

Fuzzy C-Means merupakan sebuah algoritma pengelompokkan data dimana setiap titik data milik dari sebuah *cluster* yang ditentukan oleh derajat keanggotaan.

Algoritma tersebut pertama kali dikenalkan oleh Dunn tahun 1973 dan pada tahun 1981 disempurnakan oleh Jim Bezdek (Irfan et al., 2017). Titik-titik data yang ditempatkan ke dalam kelompok berbeda dengan mempertimbangkan kesamaan nilai data dengan pusat *cluster*. *Fuzzy C-Means* bertugas untuk meminimalkan fungsi tujuan, dengan proses perubahan berulang yang melibatkan perubahan matriks derajat keanggotaan dan pusat *cluster* untuk mencapai klasifikasi yang optimal.

Pengelompokkan *fuzzy c-means* didasarkan pada teori logika *fuzzy*. Dalam teori logika *fuzzy*, keanggotaan data tidak diberi nilai pasti dengan nilai 1 (menjadi anggota) dan 0 (bukan anggota), tetapi dengan nilai derajat keanggotaan yang rentang nilainya 0 hingga 1. Nilai keanggotaan data dalam suatu himpunan adalah 0 jika data bukan anggota himpunan dan 1 jika data tersebut merupakan anggota penuh dari himpunan tersebut. Semakin tinggi nilai keanggotaan maka semakin tinggi derajat keanggotaannya, begitu pula sebaliknya (Johra, 2021).

Pada dasarnya algoritma *fuzzy c-means* memiliki banyak kesamaan dengan algoritma *K-Means.* Output dari algoritma *fuzzy c-means* bukanlah sebuah sistem inferensi *fuzzy*, melainkan deretan pusat cluster dan beberapa derajat keanggotaan untuk setiap titik pada data (Afsharizadeh et al., 2018). Berikut adalah *flowchart* dari algoritma *fuzzy c-means* pada Gambar 2.3.



Gambar 2. 3 *Flowchart Fuzzy C-Means*
Sumber : (Nurjanah, Andi Farmadi, 2014)

Tahapan algoritma *Fuzzy C-Means* dapat diuraikan sebagai berikut*:*

1) Memasukkan data keanggotaan dalam bentuk *cluster* (X), yang disusun dalam matriks berukuran *n x m* (n mewakili jumlah sampel data, m adalah atribut data) dengan X_{ij} sebagai data sampel ke-i (i = 1, 2, ...n),

atribut ke-j (*j* = 1, 2, ... *n*). Nilai keanggotaannya diperoleh dari frekuensi data dalam setiap kalimat.

2) Kemudian menentukan batasan nilai inisisalisasi untuk perhitungan sebagai berikut:

   Banyak *cluster* = *c* (disesuaikan dengan tingkat ringkasan);

   Pangkat = *w*;

   Maksimum Iterasi = *MaksIter*;

   Threshold Error = Ɛ;

   Nilai fungsi Objektif = $P_0 = 0$;

   Awal Iterasi = *t* = 1;

3) Hasilkan angka acak yang membentuk elemen matriks partisi awal U μ*ik* = 1, 2, ..., *n* ; *k* = 1,2,.....,*c*; lalu hitung total dari masing-masing kolom dengan persamaan (2.10) berikut:

$$Q_i \sum_{k=1}^{c} \mu_{ik} \qquad (2.10)$$

Keterangan :
μ_{ik} : derajat keanggotaan
Q_i : jumlah nilai derajat keanggotaan perkolom = 1 dengan *i* = 1, 2, ..., *n* ;

4) Kemudian menghitung nilai matriks partisi awal, menggunakan persamaan (2.11) berikut :

$$\mu_{ik} \frac{\mu_{ik}}{Q_i} \qquad (2.11)$$

5) Menghitung pusat *cluster* ke-k : V_{kj}, dimana *k* = 1, 2, ...,*c*; dan *j* = 1, 2, ..., *n*; dengan persamaan (2.12) berikut:

$$V_{kj} = \frac{\sum_{i=1}^{n} ((\mu_{ik})^{w} * X_{ij}}{\sum_{i=1}^{n} ((\mu_{ik})^{w}} \quad (2.12)$$

Keterangan :
V : pusat *cluster*
X_i : parameter ke-i

6) Kemudian, melakukan perhitungan nilai fungsi objektif pada iterasi ke-t (*Pt*), dengan persamaan (2.13) berikut:

$$P_t = \sum_{i=1}^{n}\sum_{k=1}^{c}\left(\left[\sum_{j=1}^{m}\left(X_{ij} - V_{kj}\right)^2\right]\mu_{ik}^{w}\right) \quad (2.13)$$

Keterangan :
Pt : nilai fungsi objektif iterasi ke-t

7) Menghitung perubahan matriks partisi U, dengan persamaan (2.14) berikut:

$$\mu_{ik} = \frac{\left[\sum_{j=1}^{m}\left(X_{ij} - V_{kj}\right)^2\right]^{\frac{-1}{w-1}}}{\sum_{k=1}^{c}\left[\sum_{j=1}^{m}\left(X_{ij} - V_{kj}\right)^2\right]^{\frac{-1}{w-1}}} \quad (2.14)$$

Keterangan :
i = 1, 2, ..., *n*;
k = 1,2, ..., *c*;

8) Terakhir melihat iterasi terakhir, jika nilai fungsi tujuan lebih kecil dari nilai ambang batas error *(|Pt – Pt--1| < ε)* atau jumlah iterasi telah melewati batas maksimum iterasi *(t* > MaksIter*)* maka berhenti. Dan jika kondisi diatas tidak terpenuhi, *t* = *t* + 1 ulangi langkah 4.

Hasil dari tahap *preprocessing* akan digunakan untuk membentuk *cluster* yang berisi kalimat-kalimat yang memiliki tingkat kemiripan yang tinggi. Setelah *cluster* atau iterasi terakhir telah terpenuhi dan didapatkan maka langkah selanjutnya memilih kalimat yang akan dijadikan ringkasan dengan menghitung

skor setiap kalimat berdasarkan nilai TF-IDF dan keanggotaan dalam *cluster*. Proses pengelompokkan ini diperlukan untuk mengorganisasikan kumpulan kalimat yang luas.

2.5 ROUGE

Matriks *Recall-Oriented Understudy for Gisting Evaluation* (ROUGE) merupakan salah satu metode evaluasi yang banyak digunakan untuk menilai kualitas hasil peringkasan teks. Metode ini bekerja dengan membandingkan hasil ringkasan sistem secara otomatis dan ringkasan referensi atau manual yang dibuat pakar. Semakin tinggi tingkat kesamaan antara keduanya, maka semakin tinggi nilai ROUGE yang diperoleh, yang menunjukkan bahwa ringkasan tersebut semakin mendekati kualitas ringkasan buatan manusia (Verma et al., 2019). Untuk mendapatkan hasil ringkasan menggunakan teknik ROUGE menggunakan persamaan (2.15) berikut:

$$ROUGE - N = \frac{\sum_{S \in \{R\}} \sum_{gram_n \in S} Count_{match}(gram_n)}{\sum_{S \in \{R\}} \sum_{gram_n \in S} Count(gram_n)} \quad (2.15)$$

Keterangan:
$Count_{march}(gram_n)$: jumlah n-gram yang cocok antara ringkasan sistem dan ringkasan manusia;
$Count(gram_n)$: jumlah total n-gram pada ringkasan manusia.

Matriks evaluasi ROUGE yang digunakan juga memiliki kelebihan dan kekurangan pada pengimplementasiannya, untuk kelebihan dari matriks ROUGE salah satu kemudahannya dalam pengimplementasian, telah menjadi standar *de facto* dalam evaluasi sistem peringkasan teks, memungkinkan perbandingan yang konsisten antar penelitian. Untuk kekurangannya sendiri yakni ketergantungan pada peringkasan referensi, jika peringkasan referensi tidak representatif atau kurang

berkualitas, maka hasil evaluasi dengan ROUGE bisa menjadi tidak akurat. Ringkasan referensi dari penelitian ini sendiri menggunakan bagian abstrak dari artikel jurnal ilmiah. ROUGE-1 membandingkan *unigram*, sedangkan ROUGE-2 membandingkan banyaknya *bigram.* Untuk menghitung jumlah *n-gram* yang sama antara teks ringkasan referensi dan ringkasa hasil sistem menggunakan ROUGE-N, dimana N menunjukan banyaknya *n-gram* yang dapat berupa 1 atau lebih dari satu. Seperti contoh berikut:

Unigrams : (saya), (suka), (makan), (kelengkeng)

Bigrams : (saya suka), (suka makan), (makan kelengkeng)

ROUGE-N merupakan sebuah pengukuran yang terkait dengan *recall* karena jumlah total dari *n-gram* yang muncul pada ringkasan referensi menjadi penyebut pada persamaannya (Lin, 2004). Nilai *precision*, *recall*, dan *f-measure* bisa dihitung menggunakan persamaan berikut:

$$Recall = \frac{\sum_{s \in sys} \sum_{gramN \in s} Count_{match}\,(gram_N)}{\sum_{s \in ref} \sum_{gramN \in s} Count(gram_N)} \qquad (2.16)$$

Nilai *recall* dapat dihitung dengan menggunakan persamaan 3.1 dimana *s* merepresentasikan kalimat atau frasa dalam ringkasan, *ref* adalah ringkasan referensi, *Count(gram N)* menunjukkan jumlah *N-gram* dalam ringkasan referensi dan $Count_{match(gramN)}$ adalah jumlah *N-gram* tertinggi yang ditemukan baik dalam ringkasan sistem maupun ringkasan referensi.

$$Precision = \frac{\sum_{s \in sys} \sum_{gramN \in s} Count_{match}\,(gram_N)}{\sum_{s \in sys} \sum_{gramN \in s} Count(gram_N)} \qquad (2.17)$$

Nilai *precision* dapat dihitung menggunakan persamaan 3.2 dimana *sys* adalah ringkasan sistem, $Count(gram_N)$ adalah jumlah N-gram yang terdapat pada ringkasan sistem. $Count_{match}(gram_N)$ merupakan jumlah maksimum N-gram yang muncul pada ringkasan sistem dan ringkasan referensi.

$$F - Score = \frac{(1 + \beta^2)(Precision * Recall)}{(\beta^2 * Precision + Recall)} \qquad (2.18)$$

$\boldsymbol{\beta}$ merupakan parameter untuk menentukan bobot relatif dari *recall* dan *precision,* dapat diatur ke 1 untuk menyeimbangkan keduanya, sehingga *F-Score* menjadi *F1-Score.* Hasil kombinasi dari nilai *recall* dan *precision* digunakan untuk mengukur kinerja dari suatu sistem.

Ringkasan sistem:

"hasil akurasi tertinggi yaitu pada kedalaman tree sebanyak 7"

Ringkasan referensi:

"nilai akurasi tertinggi didapatkan ketika kedalaman tree sebanyak 7"

Tabel 2. 1 Contoh Perhitungan ROUGE-1

Total kata unik ringkasan sistem	**Total kata unik ringkasan referensi**	**Total kata unik overlap**	***recall***	***precision***	***F1-score***
9	9	4	4/9 = 0,44	4/9 = 0,44	(2 x 0,44 x 0,44) / (0,44 + 0,44) = 0.44

Tabel 2.1 adalah contoh perhitungan skor ROUGE-1. Pada contoh perhitungan tersebut merupakan perhitungan kesamaan *unigram,* jumlah kesamaan *unigram* pada contoh kalimat tersebut sebanyak 4 kata antara lain "kedalaman" "tree" , "sebanyak", dan "7". Nilai *recall* didapatkan melalui pembagian antara jumlah kata yang sama antara kata pada ringkasan referensi dan kata pada

ringkasan sistem yaitu sebanyak 4, kemudian dibagi dengan jumlah kata pada ringkasan referensi yaitu 9, maka didapatkan nilai *recall* sebesar 0,44. Kemudian untuk nilai *precision* didapatkan dari jumlah kata yang sama antara hasil ringkasan sistem dan hasil ringkasan referensi yaitu 4 dibagi dengan jumlah seluruh kata pada ringkasan siste, yaitu 9, maka didapatkan nilai *precision* sebesar 0,44. Untuk mendapatkan skor dari *f1-score* maka harus mengkombinasikan skor dari *recall* dan *precision* yang di dapat dari persamaan 2.18, sehingga menghasilkan skor dari *f1-score* sebesar 0,44. Karena jumlah seluruh kata pada hasil ringkasan sistem dan ringkasan referensi maka hasil dari perhitungan *recall, presicion* dan *f1-score* mendapatkan hasil yang sama sebesar 0,44. *Recall* mengukur seberapa banyak elemen penting dari teks referensi yang berhasil ditangkap oleh sistem peringkasan teks. Dalam konteks ROUGE, *precision* mengukur seberapa banyak n-gram yang dihasilkan oleh sistem yang benar-benar relevan dengan teks referensi.

BAB III

DESAIN DAN IMPLEMENTASI

3.1 Pengumpulan *Dataset*

Dataset yang digunakan pada penelitian ini merupakan data hasil pengumpulan jurnal artikel ilmiah yang didapat dari berbagai sumber di laman *web* penyedia jurnal artikel ilmiah diantaranya *Google Scholar, repository* UIN Malang dan *Science and Tecnology Index* (SINTA). Jurnal yang dipilih merupakan jurnal yang berbahasa indonesia dari berbagai tema meliputi teknologi, game, sosial dan ekonomi. Artikel jurnal ilmiah merupakan hasil publikasi yang dilakukan oleh dosen, mahasiswa, ataupun seorang akademisi yang melakukan penelitian, pada penelitian ini diambil jurnal artikel ilmiah yang khusus dipublikasikan oleh mahasiswa.

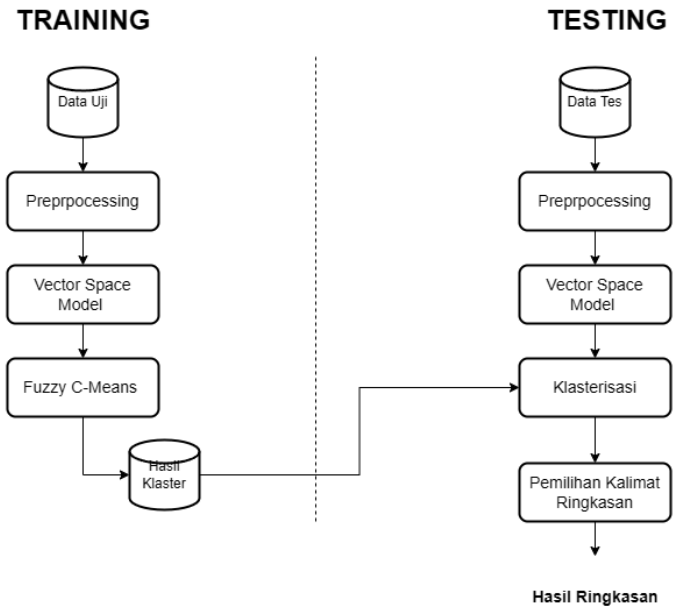
Pengumpulan *dataset* dilakukan secara manual dengan mengunduh artikel jurnal ilmiah dari laman *web open acces.* Data artikel jurnal ilmiah sebanyak 100 data yang sudah dilakukan pra-pemrosesan data, dengan pembuatan atribut judul karya ilmiah, hasil ringkasan, dan isi. Data yang digunakan berfortmat .txt pada isi artikel. Bahasa pemrograman yang akan digunakan pada penelitian ini dengan menggunakan bahasa pemrograman python.

	Judul Karya Ilmiah	Hasil Ringkasan	isi
0	Akademisi Karbitan dan Marwah Perguruan Tinggi	Pemerintah harus lebih mengevaluasi persyarata...	Jakarta - Pemerintah agaknya perlu mencermati ...
1	Disertasi Bukan Sembarang Karya	Diundang Ketua Prodi Pascasarjana di Universit...	Beberapa hari lalu saya diundang Ketua Program...
2	Membangun Tradisi Ilmiah Melalui Penelitian	Universitas pada hakikatnya sebagai institusi ...	Universitas hakikatnya adalah institusi akadem...
3	Analisis Kebutuhan Informasi Program Studi Tad...	Kebutuhan utama bagi manusia yaitu informasi. ...	Salah satu kebutuhan primer bagi manusia yaitu...
4	Membingkai Tasawuf Dengan Tafsir Ilmiah Al-Qur'an	Abdullah Darraz, menjelaskan Ayat-ayat Al-Qur'...	"Ayat-ayat al-Qur'an itu, bagaikan intan yang ...

Gambar 3. 1 Dataset Artikel Jurnal Ilmiah

3.2 Desain Sistem

Desain sistem terdapat pada Gambar 3.1 menejelaskan alur jalannya aplikasi yang akan dibuat meliputi *input* dokumen, tokenisasi, pengindeksan, pengelompokkan dengan *Fuzzy C-Means*, pembobotan dengan *Vector Space Model* dan evaluasi hasil ringkasan.



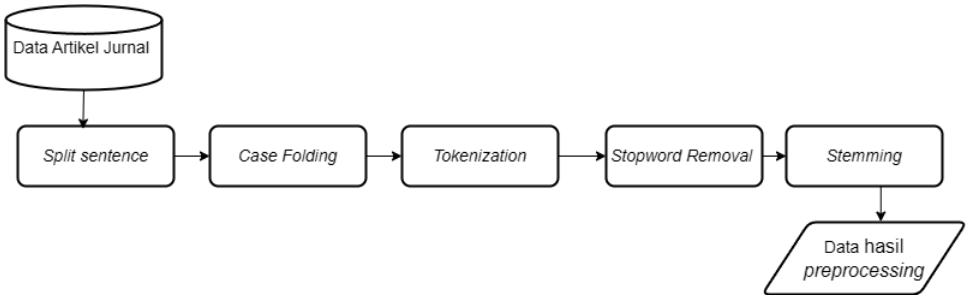
Gambar 3. 2 Desain Sistem

Langkah proses peringkasan sebagai berikut: 1. *User* memasukkan dokumen yang akan diringkas berupa artikel jurnal ilmiah berekstensi .txt. 2. Semua kalimat dalam *dataset* akan dilakukan *preprocessing* meliputi *case folding*, *stopword removal*, dan *stemming* seperti alur pada Gambar 3.2. 3. Kemudian masing-masing dari kalimat yang telah dilakukan *preprocessing* akan diberi bobot dengan perhitungan TF-IDF, setelah diberi bobot akan dibentuk sebuah vektor numerik untuk *input* pengelompokkan menggunakan *Fuzzy C-Means*. 4. Setelah token-token yang ada didalam *cluster* akan dilakukan pengelompokan berdasarkan

bobot yang telah dihitung menggunakan TF-IDF. 5. Selanjutnya data yang sudah diberi bobot akan di training model untuk *clustering* menggunakan *Fuzzy C-Means.* 6. Data hasil training akan dilakukan testing dengan dokumen baru untuk mengukur performa dan keakuratan pengelompokkan. 7. Setelah itu dilanjutkan untuk proses pemeringkatan kalimat yang sesuai dengan kelompoknya, dihitung dengan kedekatan 30% bobot tf-idf dan 70% kedekatan klaster. 8. Hasil peringkasan selanjutnya akan di evaluasi menggunakan matriks evaluasi *ROUGE-N* untuk menilai kualitas ringkasan sistem dan ringkasan manual yang dibuat pakar. 9. Setelah di evaluasi, hasil ringkasan di simpan dan dapat dilihat oleh *user* kembali.

3.3 *Preprocessing*

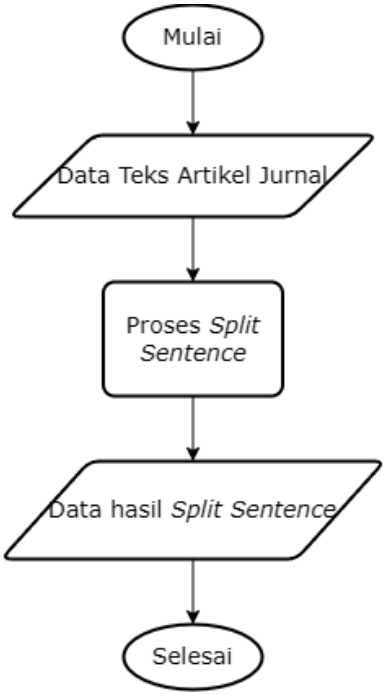
Pada tahap ini, dilakukan *preprocessing* terhadap kalimat dalam dokumen, yang mencakup pemisahan teks menjadi beberapa kalimat (*split sentence*), *case folding* untuk mengubah huruf dalam kalimat menjadi huruf kecil, *tokenazion* untuk membagi kata dalam kalimat menjadi token-token, proses *stopword removal* digunakan untuk menghailangkan kata-kata umum yang sering muncul dalam kalimat, yang tidak memiliki makna, sedangkan *stemming* berfungsi untuk mengubah setiap kata menjadi bentuk dasarnya. Kedua tahapan ini bertujuan untuk memperoleh data yang lebih bersih dan siap digunakan sebagai input dalam proses pengelompokkan serta pembobotan (Lamba & Madhusudhan, 2022). Tahapan pada *preprocessing* dapat dilihat pada Gambar 3.3.



Gambar 3. 3 Blok Diagram Tahap *Preprocessing*

3.3.1 *Split Sentence*

Split Sentence tahapan awal dalam proses *preprocessing* yaitu untuk membagi teks panjang menjadi potongan-potongan kalimat. Teks akan dipisahkan menjadi kalimat-kalimat setiap kali ditemukan titik (.) diikuti spasi atau hanya titik (.) saja. Proses ini memungkinkan untuk membuat satu paragraf menjadi potongan-potongan kalimat. Dengan memecah teks menjadi kalimat, model dapat lebih mudah mengenali struktur dan konteks, serta meningkatkan akurasi dalam prosesnya seperti penandaan bagian-bagian kalimat atau ekstraksi informasi. Alur tahap *split sentence* dijelaskan pada Gambar 3.4.



Gambar 3. 4 *Flowchart* Tahap *Split Sentence*

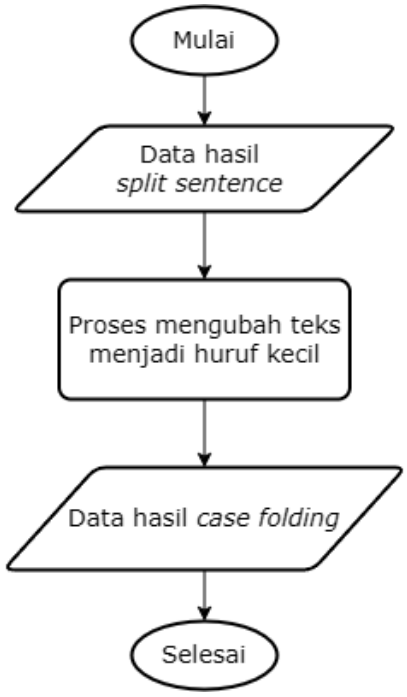
Pada tahap ini, data masukkan yang digunakan berupa teks artikel jurnal. Setiap kalimat yang diakhiri dengan tanda titik (.) akan dipisahkan menjadi satuan kalimat tersendiri. Perbandingan data sebelum dan sesudah proses *split sentence* dapat dilihat pada Tabel 3.1.

Tabel 3. 1 Contoh *Split Sentence*

Sebelum *Split Sentence*	"Indonesia merupakan salah satu negara dengan jumlah penduduk yang banyak di dunia. Dalam aktivitas sehari-hari, mobil merupakan modal transportasi yang mayoritas dipakai oleh masyarakat selain sepeda motor. Alat transportasi ini merupakan bisnis yang tergolong cepat perkembangan dan inovasinya."
Sesudah *Split Sentence*	[' Indonesia merupakan salah satu negara dengan jumlah penduduk yang banyak di dunia', 'Dalam aktivitas sehari-hari, mobil merupakanmodal transportasi yang mayoritas dipakai oleh masyarakat selain sepeda motor', 'Alat transportasi ini merupakan bisnis yang tergolong cepat perkembangan dan inovasinya',]

3.3.2 *Case Folding*

Case Folding adalah tahapan selanjutnya dari *prepocessing.* Proses ini mengubah semua huruf kapital yang ada didalam kalimat menjadi huruf kecil supaya tidak terjadi perbedaan makna yang tidak perlu didalam teks. Proses ini mengurangi kompleksitas data dan meminimalkan jumlah fitur unik yang harus diproses oleh model. Alur *case folding* dapat dilihat pada Gambar 3.5.



Gambar 3. 5 *Flowchart* Tahap *Case Folding*

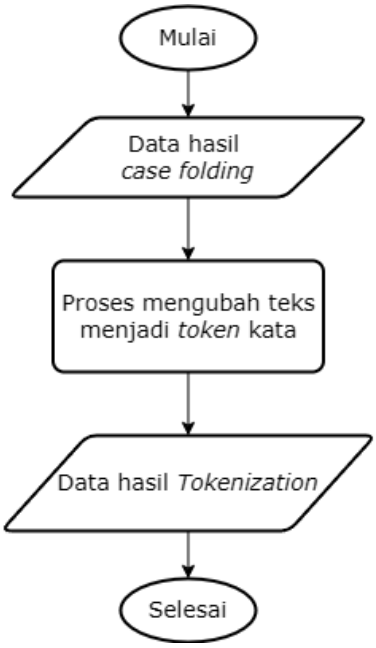
Proses ini bertujuan untuk mengurangi perbedaan makna ketika ada kata yang mengandung huruf kapital. Contoh hasil dari proses *case folding* dapat dilihat pada Tabel 3.2.

Tabel 3. 2 Contoh *Case Folding*

Sebelum *Case Folding*	[' Indonesia merupakan salah satu negara dengan jumlah penduduk yang banyak di dunia', 'Dalam aktivitas sehari-hari, mobil merupakanmodal transportasi yang mayoritas dipakai oleh masyarakat selain sepeda motor', 'Alat transportasi ini merupakan bisnis yang tergolong cepat perkembangan dan inovasinya',]
Sesudah *Case Folding*	[' indonesia merupakan salah satu negara dengan jumlah penduduk yang banyak di dunia', 'dalam aktivitas sehari-hari, mobil merupakanmodal transportasi yang mayoritas dipakai oleh masyarakat selain sepeda motor', 'alat transportasi ini merupakan bisnis yang tergolong cepat perkembangan dan inovasinya',]

3.3.3 Tokenization

Tokenization merupakan tahapan selanjutnya dalam *preprocessing.* Tahapan ini memecah kalimat menjadi kata atau token-token. Karena *dataset* yang digunakan belum dilakukan proses *tokenization* maka proses ini perlu dilakukan. *Tokenization* merupakan langkah dasar pada algoritma *Natural Language Processing* karena memungkinkan model untuk memahami teks dengan lebih terstruktur dan juga memfasilitasi proses seperti pencocokan pola, analisis frekuensi pola, analisis frekuensi kata dan pembuatan vektor kata. Alur tahap *tokenization* dapat dilihat pada Gambar 3.6.



Gambar 3. 6 *Flowchart* Tahap *Tokenization*

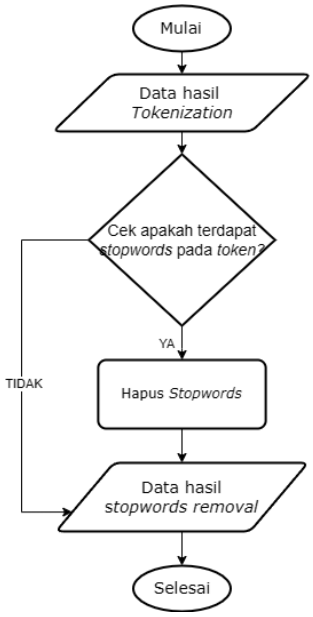
Data masukkan pada tahap ini berasal dari hasil proses *case folding*, di mana setiap kalimat diuraikan menjadi token-token kata. Perbedaan sebelum dan setelah dilakukanny tahap *tokenization* dapat dilihat pada Tabel 3.3 menunjukkan contoh proses *Tokenization.*

Tabel 3. 3 Contoh *Tokenization*

Sebelum *Tokenization*	['indonesia merupakan salah satu negara dengan jumlah penduduk yang banyak di dunia', 'dalam aktivitas sehari-hari, mobil merupakanmodal transportasi yang mayoritas dipakai oleh masyarakat selain sepeda motor', 'alat transportasi ini merupakan bisnis yang tergolong cepat perkembangan dan inovasinya',]
Sesudah *Tokenization*	[['indonesia', 'merupakan', 'salah', 'satu', 'negara', 'dengan', 'jumlah', 'penduduk', 'yang', 'banyak', 'di', 'dunia'], ['dalam', 'aktivitas', 'sehari', 'hari', 'mobil', 'merupakanmodal', 'transportasi', 'yang', 'mayoritas', 'dipakai', 'oleh', 'masyarakat', 'selain', 'sepeda', 'motor'], ['alat', 'transportasi', 'ini', 'merupakan', 'bisnis', 'yang', 'tergolong', 'cepat', 'perkembangan', 'dan', 'inovasinya'],]

3.3.4 Stopword Removal

Stopword Removal adalah sebuah proses dimana kata-kata yang tidak memberikan nilai signifikan akan dihilangkan dari sebuah teks karena kata-kata tersebut tidak memiliki nilai informasi yang relevan dalam analisis teks. Proses ini dilakukan supaya proses analisis teks berjalan dengan lebih cepat serta proses analisis yang dilakukan dapat meningkat. Contoh dari kata-kata umum yang tidak mengandung nilai yang signifikan adalah “ke”, “di”, “yang”, “dan”, dan sebagainya. Alur dari tahapan *stopword removal* dapat dilihat pada Gambar 3.7.



Gambar 3. 7 *Flowchart* Tahap *Stopword Removal*

Data masukkan yang digunakan pada tahapan ini merupakan data hasil *tokenization.* Kalimat yang telah diubah menjadi *token* kata, akan diseleksi untuk memilih kata mana yang tidak memiliki nilai yang signifikan akan dihapus. Perbedaan data sebelum dan sesudah tahap *stopword removal* dapat dilihat pada Tabel 3.4.

Tabel 3. 4 Contoh *Stopword Removal*

Sebelum *Stopword Removal*	[['indonesia', 'merupakan', 'salah', 'satu', 'negara', 'dengan', 'jumlah', 'penduduk', 'yang', 'banyak', 'di', 'dunia'], ['dalam', 'aktivitas', 'sehari', 'hari', 'mobil', 'merupakanmodal', 'transportasi', 'yang', 'mayoritas', 'dipakai', 'oleh', 'masyarakat', 'selain', 'sepeda', 'motor'], ['alat', 'transportasi', 'ini', 'merupakan', 'bisnis', 'yang', 'tergolong', 'cepat', 'perkembangan', 'dan', 'inovasinya'],]
Sesudah *Stopword Removal*	[['indonesia', 'salah', 'negara', 'penduduk', 'dunia'], ['aktivitas', 'sehari', 'mobil', 'merupakanmodal', 'transportasi', 'mayoritas', 'dipakai', 'masyarakat', 'sepeda', 'motor'], ['alat', 'transportasi', 'bisnis', 'tergolong', 'cepat', 'perkembangan', 'inovasinya'],]

Stopword Removal berfungsi untuk mempercepat proses analisis tanpa mengurangi kualitas hasil, namun harus dilakukan dengan hati-hati supaya tidak menghilangkan kata-kata yang dianggap penting.

3.3.5 *Stemming*

Stemming merupakan tahapan akhir dari *preprocessing.* Tahapan ini merupakan proses transformasi kata ke bentuk dasarnya atau akar kata dengan memanfaatkan *library* Sastrawi. Melalui proses *stemming,* kata-kata dalam teks akan memiliki bentuk representasi yang seragam sehingga lebih mudah diproses pada tahap-tahap selanjutnya. Selain itu, penggunaan *stemming* berkontribusi dalam mengurangi kompleksitas data dan mempercepat proses peringkasan. Alur dari proses *stemming* dapat dilihat pada Gambar 3.8.

Gambar 3. 8 *Flowchart* Tahap *Stemming*

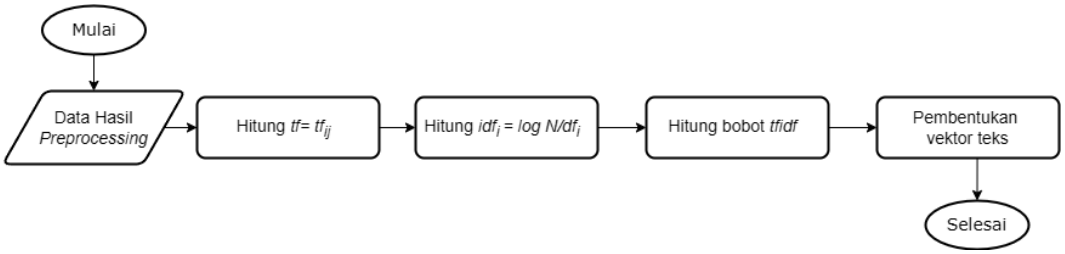
Data masukkan yang digunakan untuk tahap ini merupakan hasil dari proses *stopword removal* yang berisi *token* kata yang sudah bersih dari kata yang kurang bernilai signifikan, kemudian kata tersebut akan dipecah menjadi bentuk kata dasarnya. Perbedaan data hasil sebelum di *stemming* dan sesudah dapat dilihat pada Tabel 3.5.

Tabel 3. 5 Contoh *Stemming*

Sebelum *Stemming*	[['indonesia', 'salah', 'negara', 'penduduk', 'dunia'], ['aktivitas', 'sehari', 'mobil', 'merupakanmodal', 'transportasi', 'mayoritas', 'dipakai', 'masyarakat', 'sepeda', 'motor'], ['alat', 'transportasi', 'bisnis', 'tergolong', 'cepat', 'perkembangan', 'inovasinya'],]
Sesudah *Stemming*	[['indonesia', 'salah', 'negara', 'duduk', 'dunia'], ['aktivitas', 'hari', 'mobil', 'merupakanmodal', 'transportasi', 'mayoritas', 'pakai', 'masyarakat', 'sepeda', 'motor'], ['alat', 'transportasi', 'bisnis', 'golong', 'cepat', 'kembang', 'inovasi'],]

3.4 Implementasi *Vector Space Model*

Proses peringkasan dimulai dari pembentukkan vektor kata menggunakan *vector space model* dengan pembobotan TF-IDF sebelum dijadikan *input* untuk pengelompokkan menggunakan algoritma *Fuzzy C-Means*. Pada Gambar 3.9 menunjukkan *flowchart vector space model* pada sistem peringkasan teks secara otomatis menggunakan metode *Fuzzy C-Means* dan *Vector Space Model*.



Gambar 3. 9 *Flowchart Vector Space Model*

Pada hasil *preprocessing* dibentuk vektor kata yang akan digunakan untuk membentuk *cluster* yang berisi kalimat-kalimat yang cirinya berdekatan. Kalimat perlu direpresentasikan dalam bentuk vektor sebelum digunakan untuk proses pengelompokan dalam *cluster*. Dalam perhitungan yang dilakukan akan dihitung frekuensi setiap kata dengan dua tahapan utama, diantaranya pengukuran kemunculan jumlah kata pada dokumen (TF), dan pengukuran seberapa jauh kata tersebut muncul didalam dokumen (IDF). Proses perhitungan bobot dilakukan untuk evaluasi seberapa penting kata dalam dokumen dengan mempertimbangkan kata yang paling sering muncul dan unik pada dokumen lain yang bernilai bobot lebih tinggi. Proses perhitungan bobot menggunakan metode *Vector Space Model* sebagai berikut:

Untuk menghitung TF-IDF yang pertama kali dilakukan dengan menggabungkan kembali kalimat yang telah diproses. Dengan menggunakan persamaan (2.3) untuk menghitung nilai dari TF *(Term Frequency)* , *term frequency* merupakan banyaknya kata yang muncul dalam sebuah kalimat yang telah dilakukan *preprocessing.* Pada Tabel 3.6 menunjukkan contoh dari kalimat dalam satu dokumen artikel jurnal ilmiah.

Tabel 3. 6 Contoh kalimat pada satu dokumen

Dokumen ke-	**Kalimat hasil *preprocessing***
1	indonesia salah negara duduk dunia
2	aktivitas hari mobil modal transportasi mayoritas pakai masyarakat sepeda motor
3	alat transportasi bisnis golong cepat kembang inovasi

Selanjutnya dilakukan proses perhitungan *term frequency* untuk dokumen yang telah di *preprocessing* menggunakan rumus berikut.

$$tf_{i,j} = \frac{n_{a,b}}{\sum k \; n_{a,b}} \tag{3.1}$$

Keterangan:
$tf_{i,j}$ = Frekuensi kata
$n_{a,b}$ = Jumlah kata yang muncul dalam dokumen
$\sum k \; n_{a,b}$ = Total seluruh kata yang ada dalam dokumen

Tahapan perhitungan bobot menggunakan metode *term frequency-inverse document frequency* (tf-idf) menggunakan hasil *preprocessing* akhir yaitu *output* dari *stemming.* Pada tahap ini, teks dalam artikel jurnal akan dikonversi menjadi representasi vektor dari bobot tf-idf ditulis dengan nilai numerik, yang menghasilkan bobot setiap *term* dilihat dari seberapa sering kemunculannya serta tingkat kepentingan kata tersebut dalam dokumen.

Selanjutnya akan dicari untuk *invers document frequency* menggunakan rumus berikut.

$$idf_i = log\ (\frac{N}{dfi}) \quad (3.2)$$

Keterangan:
N = Jumlah dokumen
idf_i = Frekuensi kemunculan kata i pada semua dokumen
dfi = Dokumen yang terdapat i

Proses selanjutnya akan dihitung nilai Tf-Idf atau bobot tf-idf dengan menggunakan rumus sebagai berikut.

$$W_{ij} = tf_{ij} \,.\, idf_i \quad (3.3)$$

Keterangan:
W_{ij} = Bobot kata dalam suatu dokumen

Dimana *W* merupakan bobot dokumen yang dihasilkan dari perkalian antara *term frequency* dan *invers document frequency.*

Tabel 3.7 Menunjukkan hasil perhitungan bobot TF-IDF.

Tabel 3. 7 Perhitungan TF-IDF

Kata	***J1***	***J2***	***J3***
indonesia	0.0894	0	0
salah	0.0894	0	0
negara	0.0894	0	0
duduk	0.0894	0	0
dunia	0.0894	0	0
aktivitas	0	0.0447	0
hari	0	0.0447	0
mobil	0	0.0447	0
modal	0	0.0447	0
transportasi	0	0.0176	0.0251
mayoritas	0	0.0447	0
pakai	0	0.0447	0
masyarakat	0	0.0447	0
sepeda	0	0.0447	0
motor	0	0.0447	0
alat	0	0	0.0638
bisnis	0	0	0.0638
golong	0	0	0.0638
cepat	0	0	0.0638
kembang	0	0	0.0638

Setiap dokumen yang telah memiliki bobot, selanjutnya akan direpresentasikan sebagai vektor fitur berdasarkan bobot TF-IDF.

a. Dokumen 1 : [0.0894, 0.0894, 0.0894, 0.0894, 0.0894, 0, 0, 0, 0, ...]

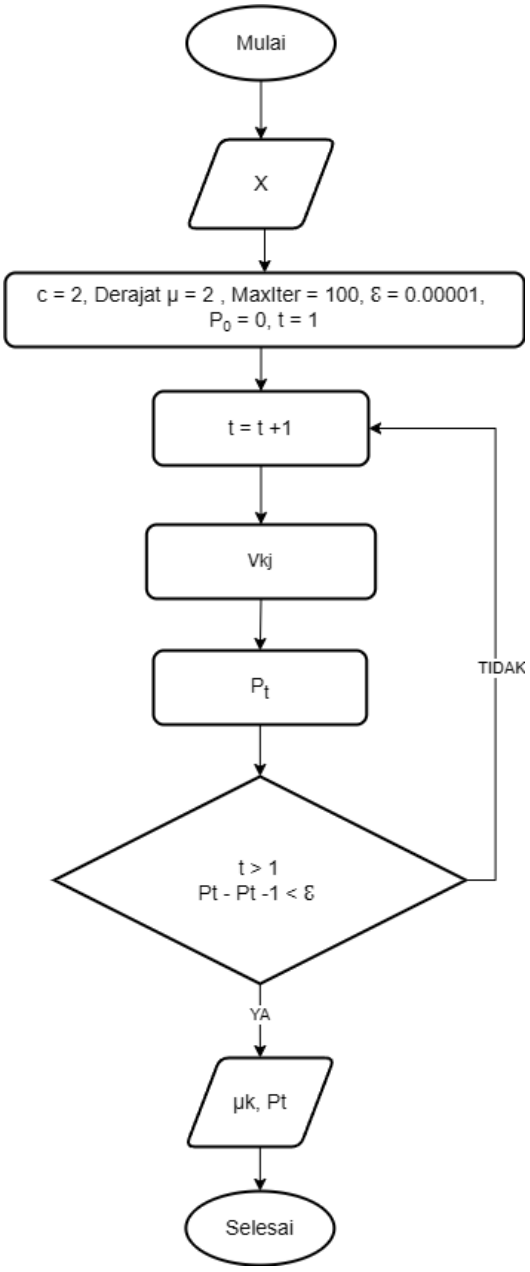
b. Dokumen 2 : [0, 0, 0, 0, 0.0477, 0.0477, 0.0477, 0.0477, 0.0176, ...]

c. Dokumen 3 : [0, 0, 0, 0, 0, 0, 0, 0, 0, 0.0251, 0, ..., 0.0638, 0.0638]

Vektor fitur ini direpresentasikan sebagai matriks berdimensi (jumlah dokumen dikali jumlah kata unik), kolom menunjukkan dokumen, dan baris menunjukkan kata-kata unik dalam korpus. Setiap baris merepresentasikan sebuah dokumen dalam ruang vektor. Matriks ini digunakan sebagai input untuk algoritma *clustering Fuzzy C-Means.*

3.5 Implementasi *Fuzzy C-Means*

Pada penelitian ini FCM digunakan untuk mengelompokkan kalimat berdasarkan bobot Tf-idf karena algoritma tersebut dalam proses prengelompokkannya menggunakan data yang berbentuk numerik. Tahapan ini dilakukan untuk mengatur proses pengumpulan data dengan cara mempartisi kumpulan data secara otomatis, sehingga objek-objek yang mempunyai persamaan akan dikelompokkan dalam suatu kelompok yang berbeda dengan kelompok lainnya (Phu et al., 2017). Gambar 3.10 merupakan *flowchart* dari *fuzzy c-means.*



Gambar 3. 10 Flowchart Fuzzy C-Means

Pembentukkan *cluster* menggunakan *fuzzy c-means* memiliki langkah-langkah sebagai berikut :

1) Data hasil matriks pembobotan TF-IDF

Data input merupakan hasil metriks pembobotan TF-IDF berupa matriks dengan dimensi (jumlah dokumen x jumlah kata unik). Data akan dimasukkan untuk proses pembentukan *cluster* menggunakan *fuzzy c-means.* Tabel 3.8. Menunjukkan vektor fitur dalam bentuk mariks TF-IDF.

Tabel 3. 8 Vektor fitur matriks TF-IDF

Kata	**Indonesia**	**salah**	**negara**	**duduk**	**dunia**	**aktivitas**	**hari**	**mobil**	**transportasi**	**.....**
Kalimat 1	0.0894	0.0894	0.0894	0.0894	0.0894	0.00	0.00	0.00	0.00	....
Kalimat 2	0.00	0.00	0.00	0.00	0.00	0.0447	0.0447	0.0447	0.0176	....
Kalimat 3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.0251	....

Pada tabel tersebut setiap nilai mewakili bobobt TF-IDF dari suatu kata dalam dokumen (kalimat ke-).

2) Inisialisasi parameter *Fuzzy C-Means*

Menginisialisasi setiap parameter yang akan digunakan :

a. Jumlah klaster $c = 2$ (misalkan)
b. Derajat μ dari (w) = 2
c. Maksimal iterasi *MaxIter* = 100
d. Threshold error $\xi = 0.00001$
e. Nilai awal fungsi objektif $P_0 = 0$
f. Iterasi awal $t = 1$

Inisialisasi ini sesuai dengan penelitian yang pernah dilakukan oleh (Suwija Putra et al., 2021) , untuk parameter memiliki syarat pengisian, parameter jumlah klaster (c) nilai lebih dari 1 dan kurang dari jumlah data sampel ($1 <$

$c < n$), parameter derajat dari μ dari (*w*) nilai lebih dari 1, iterasi maksimum (*MaxIter*) nilai lebih dari 1 dan error terkecil yang diharapkan sebesar ξ.

3) Hitung Pusat *Cluster (Centroid)*

Pada data keanggotaan awal diinisialisasi secara acak dalam bentuk matriks *U* dengan ukuran *n x c* (*n* = jumlah dokumen, *c* = jumlah cluster). Tahapan ini akan menghasilkan matriks keanggotaan awal *U* yang diinisialisasi secara acak dan dinormalisasi. Akan dibentuk dengan total 2 klaster yang akan dibuat matriks *U,* setiap elemen dalam matriks *U* memiliki nilai acak antara 0 dan 1 dan proses normalisasi, untuk membantu memastikan bahwa setiap data memiliki total keanggotaan 1. Contoh inisialisasi metriks keanggotaan untuk 3 dokumen dengan 2 klaster (*c* = 2):

Tabel 3. 9 Inisialisasi metriks *U* nilai acak

Dokumen	Cluster 1	Cluster 2
Kalimat 1	0.75	0.45
Kalimat 2	0.20	0.80
Kalimat 3	0.55	0.65

Kemudian akan dinormalisasi setiap baris agar jumlahnya = 1 menggunakan persamaan berikut

$$\mu_{ik} = \frac{\mu_{ik}}{Q_i} \tag{3.4}$$

Keterangan:
μ_{ik} = Nilai keanggotaan setelah normalisasi
Q_i = Total derajat keanggotaan sebelum normalisasi

Tabel 3.10 menunjukkan hasil dari normalisasi yang telah dilakukan.

Tabel 3. 10 Inisialisasi matriks U setelah normalisasi

Dokumen	Cluster 1	Cluster 2
Kalimat 1	0.625	0.375
Kalimat 2	0.200	0.800
Kalimat 3	0.458	0.542

Untuk menghitung pusat klaster menggunakan persamaan berikut

$$V_{kj} = \frac{\sum_{i=1}^{n} ((\mu_{ik})^{w}.\ X_{ij})}{\sum_{i=1}^{n} ((\mu_{ik})^{w}} \quad (3.5)$$

Keterangan:
V_{kj} = Pusat *cluster ke- k*
n = Jumlah total dokumen

centroid akan dihitung ulang pada setiap iterasi untuk memastikan pembagian data ke dalam *cluster* semakin akurat.

4) Perbarui *Mambership* Matrix U

Sebelum memperbarui *mambership* matrix U adalah dengan menghitung fungsi objektif iterasi ke – *t* untuk menilai perubahan fungsi objektif dibandingkan iterasi sebelumnya, menggunakan persamaan berikut.

$$P_t = \sum_{i=1}^{n}\sum_{k=1}^{c}\left(\left[\sum_{j=1}^{m}\left(X_{ij} - V_{kj}\right)^2\right]\mu_{ik}^{w}\right) \quad (3.6)$$

Keterangan:
P_t = Nilai fungsi objektif pada iterasi ke – *t*
n = Jumlah data
c = Jumlah klaster yang dibentuk
m = Dimensi atau jumlah fitur dari setiap data
X_{ij} = Nilai fitur ke-j dari data ke-i
$\left(X_{ij} - V_{kj}\right)^2$ = Jarak antara data X_{ij} dan pusat klaster V_{kj} pada dimensi *j*
$\sum_{j=1}^{m}\left(X_{ij} - V_{kj}\right)^2$ = Jarak total antara data ke-i dan pusat klaster ke-k
μ_{ik}^{w} = Derajat keanggotaan yang telah dipangkatkan dengan *w*

Persamaan diatas digunakan untuk menjumlahkan setiap data dengan pusat klasternya, dikalikan dengan derajat keanggotaan yang telah dipangkatkan. Tujuannya untuk mengevaluasi kualitas klasterisasi. Semakin kecil nilai P_t, semakin baik klasterisasi yang diperoleh. Selanjutnya untuk menghitung derajat keanggotaan yang diperbarui menggunakan persamaan berikut.

$$\mu_{ik} = \frac{\left[\sum_{j=1}^{m}\left(X_{ij} - V_{kj}\right)^2\right]^{\frac{-1}{w^{-1}}}}{\sum_{k=1}^{c}\left[\sum_{j=1}^{m}\left(X_{ij} - V_{kj}\right)^2\right]^{\frac{-1}{w^{-1}}}} \quad (3.7)$$

Keterangan:

μ_{ik} = Derajat keanggotaan data ke-i dalam klaster ke-k(antara 0 dan 1)
X_{ij} = Nilai fitur ke-j dari data ke-i
c = Jumlah klaster yang dibentuk
m = dimensi atau jumlah fitur dari setiap data

Cara kerja dari persamaan diatas dengan menghitung jarak data X_i ke pusat klaster V_k di semua dimensi dengan jarak Euclidean kuadrat, kemudian menghitung nilai kebalikan dari jarak dengan pangkat – 1/*w-1*, terakhir normalisasi nilai keanggotaan yang baru terhadap semua klaster.

5) Periksa Konvergensi

Untuk mengecek perubahan pada fungsi objektif nilai $|P_t - P_{t-1}| < \varepsilon$, jika perbedaan fungsi objektif kurang dari nilai ambang batas error maka iterasi berhenti. Jika tidak, maka iterasi dilanjutkan hingga *MaxIter* terpenuhi.

6) Hasil *Cluster*

Hasil klaster dengan keanggotaan maksimum yang akan digunakan untuk perhitungan skor kedekatan dengan klaster 70% dan kedekatan dengan bobot tf-idf sebesar 30%. Tabel 3.11 Menunjukkan hasil dari klasterisasi.

Tabel 3. 11 Hasil klasterisasi

Dokumen	**Cluster 1**	**Cluster 2**	**Keanggotaan Maksimum**
Kalimat 1	0.625	0.375	0.625
Kalimat 2	0.200	0.800	0.800
Kalimat 3	0.458	0.542	0.524

Proses peringkasan selanjutnya dengan menghitung skor dari kedekatan dengan klaster sebesar 70% dan bobot tfidf 30%. Untuk menghasilkan skor tertinggi dan hasil ringkasan akan diurutkan berdasarkan hasil skor tersebut.

3.6 Evaluasi

Hasil ringkasan pada penelitian ini akan di evaluasi menggunakan nilai ROUGE *(Recall-Oriented Understudy for Gisting Evaluation).* Nilai ROUGE merupakan matriks yang digunakan untuk mengevaluasi sistem peringkasan teks dan juga model terjemahan lain. Metriks ROUGE ini akan menghitung kata yang tumpang tindih antara ringkasan referensi dan ringkasan yang dibuat oleh sistem dengan bobot masing-masing (Zamzam, 2020). Hasil ringkasan yang sudah di urutkan pada Tabel 3.10 akan dimasukkan ke proses evaluasi menggunakan matriks ROUGE-N, kemudian akan dibandingkan antara hasil ringkasan yang dibuat oleh sistem dengan ringkasan yang dibuat manual, ringkasan manual menggunakan bagian abstrak dari artikel jurnal ilmiah. Seperti pada Tabel 3.12.

Tabel 3. 12 Contoh perbandingan hasil ringkasan

Hasil Ringkasan Sistem	Ringkasan Manual
root merupakan node yang memiliki nilai gain paling tinggi sehingga sangat mempengaruhi untuk memastikan tujuan yang akan dicapai.dalam decission tree terdapat beberapa algoritma seperti id3, c4.5, c5.0, dan cart. dari segala hasil pengujian didapatkan rata-rata akurasi dari keempat algoritma yang diusulkan berturut-turut merupakan 66,48%, 67,49%, 80,62%, serta 86,90%. decision tree adalah salah satu teknik pembelajaran mesin (machine learning) yang menggunakan hirarki aturan klasifikasi struktur berurutan dengan mempartisi dataset pelatihan secara rekursif. dari pengujian tersebut dapat ditarik kesimpulan hasil dari uji coba ini mampu memberikan hasil akurasi tertinggi yaitu pada kedalaman tree sebanyak 7 dengan akurasi 85% dan rata-rata precision 87% dan dari berbagai batasan kedalaman tree yang digunakan, dihasilkan nilai recall, precision, f1-measure, dan accuracy di atas 70%. pada penelitian ini berfokus untuk mengukur performa decision tree untuk menentukan merk mobil berdasarkan data latih dan data uji, untuk pengujiannya akan menghasilkan nilai rata-rata precision, recall, dan f1-measure.	Salah satu negara di dunia dengan jumlah penduduk yang cukup besar adalah Indonesia. Dalam aktivitas sehari-hari, mobil merupakan modal transportasi yang mayoritas dipakai oleh masyarakat selain sepeda motor. Saat ini merek mobil yang diproduksi di Indonesia maupun langsung diimpor dari luar negeri semakin banyak dengan berbagai keunggulan masing – masing. Hal tersebut menyebabkan pembeli yang mau membeli mobil seringkali kesulitan untuk menemukan merek mobil yang memenuhi spesifikasi yang diinginkan. Oleh karena itu, penelitian akan dilakukan untuk mengembangkan sistem klasifikasi merek mobil berdasarkan permasalahan di atas. Keefektifan memilih merek mobil yang akan dibeli diharapkan akan meningkat ketika menggunakan sistem ini. Dalam penelitian ini diusulkan metode decision tree. Metode decision tree sering digunakan untuk menghasilkan sebuah pohon keputusan yang memetakan kondisi dan tindakan yang harus diambil berdasarkan kondisi. Data yang digunakan memiliki Informasi tentang 3 merek mobil yakni AS, Jepang, Eropa. Tujuannya yaitu menemukan merek mobil menggunakan parameter seperti inci kubik, tahun pembuatan, dan lain-lain. Dari hasil pengujian, nilai akurasi tertinggi didapatkan ketika kedalaman tree sebanyak 7, yaitu 85%, rata-rata precision 87%, recall 80%, dan F1-Measure 83%.

Setelah dimasukkan akan dihitung menggunakan persamaan 2.15 untuk menghitung jumlah n-gram yang cocok antara ringkasan sistem dengan ringkasan manual. Setelah mengetahui hasil dari jumlah n-gram, proses selanjutnya yakni menghitung *Recall* menggunakan persamaan 2.16, menghitung *Precission* menggunakan persamaan 2.17 dan *F-Score* dihitung menggunakan persamaan 2.18. Hasil dari perhitungan tersebut akan menjadi pembanding, apakah ringkasan yang dibuat oleh sistem lebih baik dibandingkan dengan ringkasan manual.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Skenario Uji Coba

Uji coba yang dilakukan pada 100 dokumen artikel jurnal ilmiah yang telah diunduh secara manual dari *Repository* UIN Malang, *Google Schoolar,* dan *Sciene Technology Index* (SINTA). Dataset yang telah dibuat dari 100 dokumen artikel jurnal ilmiah memiliki atribut judul karya ilmiah, hasil ringkasan, dan isi. Skenario uji coba dilakukan dari id dokumen-1 sampai pada dokumen-100 untuk menghitung nilai ROUGE-N yakni dengan ROUGE-1 dan ROUGE-2 pada hasil ringkasan sistem. Sebelum dilakukan proses peringkasan pada setiap dokumen dilakukan *preprocessing*, alur dari *preprocessing* sesuai dengan Gambar 3.2 dan uji coba dilakukan dengan *preprocessing* menggunakan *Stemming* dan tidak menggunakan *Stemming*. Setelah dilakukan pra-premprosesan akan dilakukan inisialisasi parameter yang akan digunakan pada metode *Fuzzy C-Means* dengan jumlah *cluster* random sesuai hasil yang ditentukan terbaik antara 2-6 *cluster*, parameter *fuzziness* f 2, max_interasi 100, dan nilai *epsilon* 0,00001. Proses menghitung kesamaan kosinus antar kalimat dengan vektor dokumen dan menghitung skor akhir untuk setiap kalimat dengan kedekatan *cluster* menggunakan metode *Vector Space Model.* Panjang ringkasan yang dihasilkan sistem sebesar 10%, 20%, 30%, 40% dan 50% dari keseluruhan kalimat pada teks asli.

4.2 Hasil Uji Coba

Berdasarkan pada skenario uji coba yang dijabarkan pada subbab 4.1 pengujian yang telah dilakukan untuk perbandingan ringkasan manual dengan ringkasan oleh sistem yang telah dibuat. Penelitian ini menggunakan evaluasi ROUGE-N yakni menggunakan ROUGE-1 dan ROUGE-2 dalam mendapatkan nilai *recall, pressicion, f-measure* dan *accuracy.* Tabel 4.1 merupakan hasil dari tahap preprocessing dari setiap dokumen yang ada dalam dataset.

Tabel 4. 1 Hasil *preprocessing*

Index Dokumen	Isi Dokumen	Hasil Preprocessing
0	Syntax Jurnal Informatika merupakan sebuah sis.....	syntax jurnal informatika rupa buah sistem inf...
1	Hati atau liver adalah organ tubuh yang terlet.....	hati liver organ tubuh letak bagi atas rongga ...
2	Secara geografis, Indonesia adalah salah satu.....	geografis indonesia salah satu negara pulau be...
3	Salah ssatu permasalahan serius yang dihadapi a.....	salah satu masalah serius hadap akibat tanam t...
4	Indonesia merupakan salah satu negara dengan j.....	indonesia rupa salah satu negara jumlah duduk .....
.....	.....	.....
95	Penumpukkan sampah yang tidak sesuai dengan kap.....	tumpu sampah sesuai kapasitas masalah nyata.....
96	Secara umum, Diabetes dikenal sebagai penyakit.....	diabetes kenal sakit epidemi dampak negara kelompok.....
97	Pencemaran udara menjadi permasalahan yang ser.....	cemar udara masalah seringkali jumpa....
98	Kartu Kredit adalah sebuah alat pembayaran yan.....	kartu kredit alat bayar keluar bank.....
99	Serangan Jantung adalah salah satu penyakit ya.....	serang jantung salah sakit mati dunia.....

Pada hasil *preprocessing* diatas akan digunakan untuk input ke dalam model yaitu pada proses *clustering* dan perhitungan *Term Frequency Invers Document Frequency.*

Tabel 4. 2 Derajat keanggotaan dokumen terhadap *cluster* dengan *stemming*

Data Point	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
0	0.166571	0.166571	0.166571	0.166571	0.167147	0.166571
1	0.000545	0.000545	0.000545	0.000545	0.997276	0.000545
2	0.166618	0.166618	0.166618	0.166618	0.166912	0.166618
3	0.000545	0.000545	0.000545	0.000545	0.997276	0.000545
4	0.000545	0.000545	0.000545	0.000545	0.997276	0.000545
.....	.....	.....	.....	.....	.....	.....
48633	0.000545	0.000545	0.000545	0.000545	0.997276	0.000545
48634	0.166636	0.166636	0.166636	0.166636	0.166818	0.166636
48636	0.166646	0.166646	0.166646	0.166646	0.166771	0.166646
48636	0.166571	0.166571	0.166571	0.166571	0.167147	0.166571
48637	0.0005545	0.0005545	0.0005545	0.0005545	0.997276	0.0005545

Pada Tabel 4.2 menunjukkan nilai derajat keanggotaan semua data terhadap setiap *cluster* dengan menggunakan *preprocessing stemming,* dibentuk sebanyak 6 *cluster*. Data *point* berjumlah 48637 dihasilkan dari tfidf_matrix yang dibentuk sebelumnya, setiap kata unik dalam seluruh korpus dokumen dianggap sebagai satu fitur, dalam satu dokumen terdiri dari teks panjang dan beragam, dengan begitu kata unik yang dihasilkan ribuan. Pada tabel tersebut terdapat 48637 kata unik di antara dokumen, tfidf_matrix memiliki dimensi (100, 48637) kemudian di *transpose* maka matriksnya menjadi berdimensi (48637, 100).

Tabel 4. 3 Derajat keanggotaan dokumen terhadap *cluster* tanpa *stemming*

Data Point	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
0	0.166604	0.166604	0.166604	0.166604	0.166978	0.166604
1	0.000559	0.000559	0.000559	0.000559	0.997204	0.000559
2	0.166637	0.166637	0.166637	0.166637	0.166813	0.166637
3	0.000559	0.000559	0.000559	0.000559	0.997204	0.000559
4	0.000559	0.000559	0.000559	0.000559	0.997204	0.000559
.....	.....	.....	.....	.....	.....	.....
48633	0.000559	0.000559	0.000559	0.000559	0.997204	0.000559
48634	0.166604	0.166604	0.166604	0.166604	0.166978	0.166604
48636	0.166680	0.166680	0.166680	0.166680	0.166602	0.166680
48636	0.166604	0.166604	0.166604	0.166604	0.166978	0.166604
48637	0.000559	0.000559	0.000559	0.000559	0.997204	0.000559

Pada Tabel 4.3 diatas menunjukkan nilai derajat keanggotaan semua data terhadap setiap *cluster* tanpa menggunakan *preprocessing stemming,* dibentuk

sebanyak 6 *cluster*. Data *point* berjumlah 48637 dihasilkan dari tfidf_matrix yang dibentuk sebelumnya, setiap kata unik dalam seluruh korpus dokumen dianggap sebagai satu fitur, dalam satu dokumen terdiri dari teks panjang dan beragam, dengan begitu kata unik yang dihasilkan ribuan. Pada tabel tersebut terdapat 48637 kata unik di antara dokumen, tfidf_matrix memiliki dimensi (100, 48637) kemudian di transpose, maka matriksnya menjadi berdimensi (48637, 100). Selanjutnya untuk hasil dari perhitungan kedekatan *cluster* sebesar 30% dan Tf-Idf sebesar 70% yang menggunakan *stemming* dapat dilihat pada Tabel 4.4.

Tabel 4. 4 Hasil perhitungan kedekatan *cluster* dan TF-IDF dengan *stemming*

	Kalimat ke -	**skor**
Dokumen 1	2	4.1199
	4	4.0949
	5	4.0899
	7	4.0842
	12	4.0783
Dokumen 2	2	3.8720
	4	3.8470
	5	3.8420
	7	3.8363
	12	3.8303
Dokumen 3	2	3.4790
	4	3.4540
	5	3.4490
	7	3.4433
	12	3.4374
Dokumen 4	2	3.4054
	4	3.3804
	5	3.3754
	7	3.3697
	12	3.3637

Tabel 4. 5 Hasil perhitungan kedekatan *cluster* dan TF-IDF tanpa *stemming*

	Kalimat ke -	**skor**
Dokumen 1	2	4.4522
	4	4.4272
	5	4.4222
	7	4.4164
	12	4.4105
Dokumen 2	2	4.4132
	4	4.3882
	5	4.3832
	7	4.3775
	12	4.3715
Dokumen 3	2	3.7960
	4	3.7710
	5	3.7660
	7	3.7603
	12	3.7544
Dokumen 4	2	3.5702
	4	3.5452
	5	3.5402
	7	3.5345
	12	3.5286

Dari kedua tabel diatas merupakan hasil dari urutan skor tertinggi dari setiap kalimat yang berada didalam dokumen dengan 6 *cluster*, tabel tersebut menampilkan lima kalimat dengan skor tertinggi yang berada di setiap dokumen yang menggunakan *stemming* dan dokumen tanpa menggunakan *stemming.*

Dapat dilihat pada tabel urutan kalimat yang dihasilkan sama, karena indeks kalimat diurutkan setelah memilih kalimat dengan skor tertinggi dan untuk hasil skor berbeda, karena pengaruh dari proses *stemming* dan tanpa *stemming,* yang menghasilkan dimensi matriks tf-idf yang berbeda. Kemudian dari hasil tersebut nantinya akan diurutkan berdasarkan skor tertinggi dan dimasukkan ke dalam ringkasan pada tingkat kompresi 10%, 20%, 30%, 40%, dan 50%.

4.2.1 Tingkat Peringkasan 10%

Pada uji coba pertama untuk tingkat peringkasan sebesar 10% dari keseluruhan isi dokumen. *Cluster* optimal pada 6 *cluster*, parameter *fuzziness f* 2, dan max_iterasi 100. Pada Tabel 4.4 dan Tabel 4.5 bisa dilihat hasil clustering dari 6 jumlah *cluster* yang sudah dianggap sebagai *cluster* optimal oleh model.

Jumlah kalimat yang sesuai dengan ringkasan manual sangat mempengaruhi hasil evaluasi dari nilai ROUGE-1 dan ROUGE-2 dengan menghitung *precision, recall,* dan *f-score* pada setiap dokumen artikel jurnal ilmiah. Hasil rata-rata evaluasi ringkasan untuk *Compression rate* 10% ditunjukkan pada Tabel 4.6 dan Tabel 4.7.

Tabel 4. 6 Hasil rata-rata ROUGE *Compression Rate* 10% dengan *Stemming*

Compression Rate	**ROUGE-N**							
	ROUGE-1				**ROUGE-2**			
	Precision	***Recall***	***F1-Score***	***Accuracy***	***Precision***	***Recall***	***F2-Score***	***Accuracy***
10%	0.2321	0.6278	0.3271	0,4300	0.0539	0.1414	0.0751	0,0976

Tabel 4. 7 Hasil rata-rata ROUGE *Compression Rate* 10% tanpa *Stemming*

Compression Rate	**ROUGE-N**							
	ROUGE-1				**ROUGE-2**			
	Precision	***Recall***	***F1-Score***	***Accuracy***	***Precision***	***Recall***	***F2-Score***	***Accuracy***
10%	0.2321	0.6278	0.3271	0,4300	0.0539	0.1414	0.0751	0,0976

Contoh dari hasil ringkasan manual, ringkasan sistem menggunakan *Stemming*, dan tanpa *Stemming* pada dokumen id-5 yang berjudul “Rekomendasi Merk Mobil Untuk Calon Pembeli Menggunakan Algoritma Decision Tree” dapat dilihat pada Gambar 4.1, Gambar 4.2 dan Gambar 4.3.

Salah satu negara di dunia dengan jumlah penduduk yang cukup besar adalah Indonesia. Dalam aktivitas sehari-hari, mobil merupakan modal transportasi yang mayoritas dipakai oleh masyarakat selain sepeda motor. Saat ini merek mobil yang diproduksi di Indonesia maupun langsung diimpor dari luar negeri semakin banyak dengan berbagai keunggulan masing – masing. Hal tersebut menyebabkan pembeli yang mau membeli mobil seringkali kesulitan untuk menemukan merek mobil yang memenuhi spesifikasi yang diinginkan. Oleh karena itu, penelitian akan dilakukan untuk mengembangkan sistem klasifikasi merek mobil berdasarkan permasalahan di atas. Keefektifan memilih merek mobil yang akan dibeli diharapkan akan meningkat ketika menggunakan sistem ini. Dalam penelitian ini diusulkan metode decision tree. Metode decision tree sering digunakan untuk menghasilkan sebuah pohon keputusan yang memetakan kondisi dan tindakan yang harus diambil berdasarkan kondisi. Data yang digunakan memiliki Informasi tentang 3 merek mobil yakni AS, Jepang, Eropa. Tujuannya yaitu menemukan merek mobil menggunakan parameter seperti inci kubik, tahun pembuatan, dan lain-lain. Dari hasil pengujian, nilai akurasi tertinggi didapatkan ketika kedalaman tree sebanyak 7, yaitu 85%, rata-rata precision 87%, recall 80%, dan F1-Measure 83%.

Gambar 4. 1 Ringkasan Manual id-5

agar dapat mempertahankan penjualan di tengah banyaknya barang yang bersaing, para perusahaan pun berlomba-lomba mempelajari faktor-faktor apa saja yang mempengaruhi penjualan produk yang mereka buat dan memahami secara mendalam keinginan calon pembeli. decision tree adalah salah satu teknik pembelajaran mesin (machine learning) yang menggunakan hierarki aturan klasifikasi struktur berurutan dengan mempartisi untuk menghasilkan sebuah pohon keputusan yang memetakan kondisi dan tindakan yang harus diambil berdasarkan kondisi tersebut. pada penelitian terdahulu, penggunaan algoritma decision tree pernah dilakukan untuk judul penelitian "rekomendasi pengambilan mata kuliah pilihan untuk mahasiswa sistem informasi menggunakan algoritme decision tree ". metode yang diusulkan dalam penelitian tersebut menggunakan metode decision tree untuk menghasilkan rekomendasi mata kuliah pilihan berdasarkan sejumlah kriteria yang relevan. penelitian terkait lainnya berjudul "pemilihan jenis smartphone sesuai dengan kebutuhan menggunakan metode forward chaining dan decision tree ". pengujian pada penelitian rujukan berfokus pada pengujian aplikasi yang dihasilkan, dengan menggunakan black box testing , yaitu pengujian yang dilakukan dengan cara mengamati input dari pengguna dan output dari hasil aplikasinya. pada penelitian ini berfokus untuk mengukur performa decision tree untuk menentukan merek mobil berdasarkan data latih dan data uji, untuk pengujiannya akan menghasilkan nilai rata-rata precision , recall , dan f1-measure. pada decision tree juga ada istilah asm atau attribute selection measure , yaitu cara yang dapat digunakan untuk membantu menemukan kriteria pemisah yang optimal untuk mengelompokkan data. dari pengujian tersebut, dapat ditarik kesimpulan hasil uji coba ini mampu memberikan hasil akurasi tertinggi yaitu pada kedalaman tree sebanyak 7 dengan akurasi 85% dan rata-rata precision 87%.

Gambar 4. 2 Ringkasan *Compression Rate* 10% id-5 dengan *Stemming*

agar dapat mempertahankan penjualan di tengah banyaknya barang yang bersaing, para perusahaan pun berlomba-lomba mempelajari faktor-faktor apa saja yang mempengaruhi penjualan produk yang mereka buat dan memahami secara mendalam keinginan calon pembeli. decision tree adalah salah satu teknik pembelajaran mesin (machine learning) yang menggunakan hierarki aturan klasifikasi struktur berurutan dengan mempartisi untuk menghasilkan sebuah pohon keputusan yang memetakan kondisi dan tindakan yang harus diambil berdasarkan kondisi tersebut. pada penelitian terdahulu, penggunaan algoritma decision tree pernah dilakukan untuk judul penelitian "rekomendasi pengambilan mata kuliah pilihan untuk mahasiswa sistem informasi menggunakan algoritme decision tree ". metode yang diusulkan dalam penelitian tersebut menggunakan metode decision tree untuk menghasilkan rekomendasi mata kuliah pilihan berdasarkan sejumlah kriteria yang relevan. penelitian terkait lainnya berjudul "pemilihan jenis smartphone sesuai dengan kebutuhan menggunakan metode forward chaining dan decision tree ". pengujian pada penelitian rujukan berfokus pada pengujian aplikasi yang dihasilkan, dengan menggunakan black box testing , yaitu pengujian yang dilakukan dengan cara mengamati input dari pengguna dan output dari hasil aplikasinya. pada penelitian ini berfokus untuk mengukur performa decision tree untuk menentukan merek mobil berdasarkan data latih dan data uji, untuk pengujiannya akan menghasilkan nilai rata-rata precision , recall , dan f1-measure. pada decision tree juga ada istilah asm atau attribute selection measure , yaitu cara yang dapat digunakan untuk membantu menemukan kriteria pemisah yang optimal untuk mengelompokkan data. dari pengujian tersebut, dapat ditarik kesimpulan hasil uji coba ini mampu memberikan hasil akurasi tertinggi yaitu pada kedalaman tree sebanyak 7 dengan akurasi 85% dan rata-rata precision 87%.

Gambar 4. 3 Ringkasan *Compression Rate* 10% id-5 tanpa *Stemming*

4.2.2 Tingkat Peringkasan 20%

Pada peringkasan dengan tingkat kompresi 20% parameter yang digunakan sama seperti peringkasan tingkat kompresi 10% yang berbeda hanya tingkat kompresi peringkasannya saja. Hasil dari derajat keanggotaan juga tetap sama seperti tingkat peringkasan 10%. Tabel 4.8. menunjukkan hasil rata-rata evaluasi skor ROUGE-1 dan ROUGE-2 pada peringkasan dengan tingkat kompresi 20% dengan *Stemming* dan Tabel 4.9 tanpa *Stemming.*

Tabel 4. 8 Hasil rata-rata ROUGE *Compression Rate* 20% dengan *Stemming*

Compression Rate	**ROUGE-N**							
	ROUGE-1				**ROUGE-2**			
	Precision	***Recall***	***F1-Score***	***Accuracy***	***Precision***	***Recall***	***F2-Score***	***Accuracy***
20%	0.4053	0.5643	0.4530	0,4848	0.0984	0.1325	0.1080	0,1154

Tabel 4. 9 Hasil rata-rata ROUGE *Compression Rate* 20% tanpa *Stemming*

Compression Rate	**ROUGE-N**							
	ROUGE-1				**ROUGE-2**			
	Precision	***Recall***	***F1-Score***	***Accuracy***	***Precision***	***Recall***	***F2-Score***	***Accuracy***
20%	0.4049	0.5640	0.4526	0,4844	0.0984	0.1325	0.1080	0,1153

Contoh dari hasil ringkasan manual dan ringkasan sistem pada dokumen id-21 yang berjudul "Analisis Perbandingan Metode Alpha Miner, Inductive Miner Dan Causal-Net Mining Dalam Proses Mining" dapat dilihat pada Gambar 4.4, Gambar 4.5 dan Gambar 4.6.

Tidak semua algoritma proses mining dapat mendeteksi semua skenario model proses, sehingga dilakukan eksperimen dengan mecoba 3 jenis algoritma terhadap 9 skenario model proses bisnis yang bertujuan untuk mendapatkan algoritma yang paling cocok untuk setiap skenario model proses. Kami menggunakan 3 algoritma proses mining diantaranya alpha miner, inductive miner dan causal-net mining. Kami mengusulkan solusi dengan menggunakan aplikasi ProM untuk mengecek kecocokan 3 algoritma yang digunakan terhadap 9 skenario. Selain itu, untuk mendukung hasil proses mining menggunakan ProM, kami mengukur nilai similarity dengan membandingkan model proses pada dataset dengan hasil proses mining menggunakan ProM. Berdasarkan hasil pengukuran similarity diketahui bahwa eksperimen menggunakan algortima alpha miner. Pada figure 8 memiliki nilai tingkat similarity paling tinggi yaitu dengan nilai 0.89. Sedangkan tingkat similarity paling kecil, didapati pada figure 7 menggunakan alpha miner dengan nilai 0.12.

Gambar 4. 4 Ringkasan Manual id-21

tujuannya yaitu untuk memperkaya atau memperluas model yang sudah ada dengan informasi yang didapat dari event log di dunia nyata, dan untuk mengganti atau menambah model apriori. algoritma ini pertama kali dikemukakan oleh van der aalast, seorang professor di departemen matematika dan ilmu komputer dari technische universiteit eindhoven (tu/e). menurut weijters algoritma heuristic miner memiliki tiga langkah utama, yaitu membuat grafik dependensi, membuat tanda input dan output untuk setiap aktivitas dan mencari hubungan ketergantungan jarak jauh. secara khusus, ini berfokus pada proses discovery sebagai sarana untuk merekonstruksi struktur terkait proses dari event log, seperti aliran kontrol proses, jaringan sosial, dan aliran data. dalam makalah ini, peneliti menunjukkan bahwa analisis delta dan pengujian confermance dapat digunakan untuk menganalisis penyelarasan bisnis selama actual event dicatat dan user memiliki kendali atas proses tersebut. berdasarkan teknik proses mining seperti proses discovery dan confermance checking, perspektif yang hilang ini diharapkan akan memungkinkan penemuan pola pengkodean, pencarian perilaku programmer, dan deteksi penyimpangan dari proses yang ditentukan. pada penelitian ini, kami meggunakan discovery untuk melakukan proses mining, karena diketahui bahwa proses discovery dibangun dari bawah ke atas khusus untuk rpa itulah sebabnya proses discovery memberikan yang lebih cepat, lebih fleksibel, pendekatan yang lebih komprehensif, dan lebih hemat biaya untuk secara otomatis mengidentifikasi dan menganalisis proses kerja otomatis. nilai kesamaan pada similarity struktur diukur dengan membandingkan aspek yang serupa pada model proses bpmn, seperti label task activity, connector, dan gateway (percabangan) yang sama. dengan menggunakan event log pada tabel 1, dilakukan proses mining dengan menggunakan tiga algoritma yaitu algoritma alpha miner, inductive miner dan causal-net mining.berdasarkan hasil penelitian pada table 3 setelah kami lakukan perbandingan, diketahui bahwa setiap model proses hasil discovery memiliki nilai similarity yang beragam. pada algoritma alpha miner menghasilkan model roses dalam bentuk petri net, oleh karena itu kami mengubahnya menjadi bentuk bpmn dengan format .pnml menggunakan software wopad. selain itu, ada pula figure 9a dan 9b dengan hasil similarity yang error dikarenakan saat melakukan discovery mengunakan algoritma alpha miner, model proses yang dihasilkan terpecah menjadi 2 bagian, sehingga tidak bisa dideteksi nilai similarity nya. berdasarkan pembahasan yang telah diuraikan sebelumnya, kami menyimpulkan bahwa pada penelitian ini kami menemukan bahwa, nilai tingkat similarity tertinggi didapati pada figure 8 menggunakan alpha miner dengan nilai 0.89.

Gambar 4. 5 Ringkasan Sistem *Compression Rate* 20% id-21 dengan *Stemming*

tujuannya yaitu untuk memperkaya atau memperluas model yang sudah ada dengan informasi yang didapat dari event log di dunia nyata, dan untuk mengganti atau menambah model apriori. algoritma ini pertama kali dikemukakan oleh van der aalast, seorang professor di departemen matematika dan ilmu komputer dari technische universiteit eindhoven (tu/e). menurut weijters algoritma heuristic miner memiliki tiga langkah utama, yaitu membuat grafik dependensi, membuat tanda input dan output untuk setiap aktivitas dan mencari hubungan ketergantungan jarak jauh. secara khusus, ini berfokus pada proses discovery sebagai sarana untuk merekonstruksi struktur terkait proses dari event log, seperti aliran kontrol proses, jaringan sosial, dan aliran data. dalam makalah ini, peneliti menunjukkan bahwa analisis delta dan pengujian confermance dapat digunakan untuk menganalisis penyelarasan bisnis selama actual event dicatat dan user memiliki kendali atas proses tersebut. berdasarkan teknik proses mining seperti proses discovery dan confermance checking, perspektif yang hilang ini diharapkan akan memungkinkan penemuan pola pengkodean, pencarian perilaku programmer, dan deteksi penyimpangan dari proses yang ditentukan. pada penelitian ini, kami meggunakan discovery untuk melakukan proses mining, karena diketahui bahwa proses discovery dibangun dari bawah ke atas khusus untuk rpa itulah sebabnya proses discovery memberikan yang lebih cepat, lebih fleksibel, pendekatan yang lebih komprehensif, dan lebih hemat biaya untuk secara otomatis mengidentifikasi dan menganalisis proses kerja otomatis. nilai kesamaan pada similarity struktur diukur dengan membandingkan aspek yang serupa pada model proses bpmn, seperti label task activity, connector, dan gateway (percabangan) yang sama. dengan menggunakan event log pada tabel 1, dilakukan proses mining dengan menggunakan tiga algoritma yaitu algoritma alpha miner, inductive miner dan causal-net mining.berdasarkan hasil penelitian pada table 3 setelah kami lakukan perbandingan, diketahui bahwa setiap model proses hasil discovery memiliki nilai similarity yang beragam. pada algoritma alpha miner menghasilkan model roses dalam bentuk petri net, oleh karena itu kami mengubahnya menjadi bentuk bpmn dengan format .pnml menggunakan software wopad. selain itu, ada pula figure 9a dan 9b dengan hasil similarity yang error dikarenakan saat melakukan discovery mengunakan algoritma alpha miner, model proses yang dihasilkan terpecah menjadi 2 bagian, sehingga tidak bisa dideteksi nilai similarity nya. berdasarkan pembahasan yang telah diuraikan sebelumnya, kami menyimpulkan bahwa pada penelitian ini kami menemukan bahwa, nilai tingkat similarity tertinggi didapati pada figure 8 menggunakan alpha miner dengan nilai 0.89.

Gambar 4. 6 Ringkasan Sistem *Compression Rate* 20% id-21 tanpa *Stemming*

4.2.3 Tingkat Peringkasan 30%

Pada peringkasan dengan tingkat kompresi 30% sama seperti pada peringkasan dengan tingkat kompresi 20% menggunakan parameter yang sama dan dengan *cluster* yang sama. Hasil nilai rata-rata dari evaluasi ROUGE-N ditunjukkan pada Tabel 4.10 dengan *Stemming* dan Tabel 4.11 tanpa *Stemming.*

Tabel 4. 10 Hasil rata-rata ROUGE *Compression Rate* 30% dengan *Stemming*

Compression Rate	**ROUGE-N**							
	ROUGE-1				**ROUGE-2**			
	Precision	***Recall***	***F1-Score***	***Accuracy***	***Precision***	***Recall***	***F2-Score***	***Accuracy***
30%	0.5331	0.5034	0.4975	0,5183	0.1369	0.1243	0.1249	0,1306

Tabel 4. 11 Hasil rata-rata ROUGE *Compression Rate* 30% tanpa *Stemming*

Compression Rate	**ROUGE-N**							
	ROUGE-1				**ROUGE-2**			
	Precision	***Recall***	***F1-Score***	***Accuracy***	***Precision***	***Recall***	***F2-Score***	***Accuracy***
30%	0.5330	0.5033	0.4974	0,5181	01368	01242	0.1248	0,1305

Contoh dari hasil ringkasan manual dan ringkasan sistem pada dokumen id-14 yang berjudul "Klasifikasi Jenis Kayu Menggunakan Support Vector Machine Berdasarkan Ciri Tekstur Local Binary Pattern" dapat dilihat pada Gambar 4.7, Gambar 4.8 dan Gambar 4.9.

Indonesia merupakan negara yang kaya akan sumber daya alam, salah satunya adalah kayu. Kayu maupun produk dari kayu merupakan komoditas unggulan ekspor. Mengingat banyaknya jenis kayu yang memiliki tekstur yang hampir sama, maka diperlukan pemahaman untuk mengenalinya. Identifikasi jenis kayu saat ini pada umumnya masih dilakukan manusia secara visual. Kemampuan mengidentifikasi jenis kayu harus dilakukan secara berulang-ulang dan membutuhkan waktu proses latihan yang lama. Keterbatasan kemampuan manusia dalam mengidentifikasi jenis kayu yang belum terampil secara visual terkadang berpengaruh terhadap hasil yang diinginkan bagi dunia industri. Saat ini teknologi pengolahan citra digital telah banyak digunakan untuk melakukan klasifikasi jenis kayu berdasarkan teksturnya. Pada penelitian ini metode local binary pattern (LBP) digunakan untuk melakukan klasifikasi jenis kayu. Metode ini akan menghasilkan ciri tekstur yang akan digunakan sebagai masukan dalam proses pelatihan dan pengujian menggunakan support vector machine (SVM). Ciri tekstur yang digunakan dalam metode LBP ini adalah mean, standar deviasi, skewness, energi, dan entropi. Data citra jenis kayu yang digunakan dalam penelitian ini adalah jenis kayu bayur, cempaka, damar, meranti, dan merbau. Citra kayu tersebut diambil secara manual menggunakan kamera digital dengan jarak pengambilan 20 cm. Hasil akurasi klasifikasi terhadap citra jenis kayu bayur, cempaka, damar, meranti, dan merbau dalam penelitian ini dengan jarak ketetanggaan R=1 adalah sebesar 91,3% berada pada parameter sigma 0,3. Sedangkan hasil error terkecil hasil klasifikasi adalah sebesar 8,7%.

Gambar 4. 7 Ringkasan manual id-14

dewasa ini hutan tanaman yang menanam berbagai jenis kayu baik dari jenis unggulan setempat (native species) maupun dari jenis eksotik (exotic species) makin berkembang, contohnya hutan rakyat, hutan kemasyarakatan, dan hutan industri. keterbatasan kemampuan manusia dalam mengidentifikasi jenis kayu yang belum terampil secara visual terkadang berpengaruh terhadap hasil yang diinginkan bagi dunia industri. keuntungan dari metode lbp adalah invarian untuk perubahan skala abu-abu monotonik, kompleksitas komputasi yang rendah dan multi skala yang nyaman. hasil akhir menunjukkan bahwa system dapat mendeteksi adanya cacat pada kayu dengan tingkat akurasi tertinggi adalah 89,4%, far sebesar 7,6% dan frr sebesar 3%, dengan waktu komputasi rata-rata sistem sebesar 0,3069 detik [5]. klasifikasi jenis kayu jati, sengon, mahoni dan mindi pada penelitian terdahulu telah dilakukan menggunakan back-propagation neural network berdasarkan fitur gray level co□occurrence matrices. hasil penelitian ini menunjukkan nilai rata□rata akurasi yaitu 96.13% [6].penelitian mengenai ekstraksi citra kayu juga telah dilakukan sebelumnya pada jenis kayu jati, mahoni, mindi, dan albasia untuk menentukan fitur dan arah sudut yang lebih efisien dan efektif dalam mengidentifikasi spesies kayu. jenis fitur yang diuji dalam penelitian adalah angular second moment (asm), contrast, idm/homogenitas, entropi, korelasi dengan arah komputasi yakni 0, 45, 90, dan 135 derajat glcm. mesin pembelajaran untuk melakukan klasifikasi tekstur citra yang dapat digunakan diantaranya adalah logika fuzzy, jaringan syaraf tiruan, algoritma genetika, dan support vector machine (svm) [8]. penelitian ini menghasilkan tingkat pengenalan terbaik yakni 87,5% berada pada jarak pengambilan 20 cm dengan jarak piksel tetangga d=2 pada arah glcm 135 derajat. sedangkan penambahan kanal warna yakni hue, saturation, dan value (hsv) pada klasifikasi citra daging ini, menghasilkan klasifikasi terbaik yakni 75,6% yang berada pada arah glcm 45 derajat dengan jarak piksel tetangga d=3 dan arah glcm 135 derajat dengan jarak piksel tetangga d=2. dalam penelitian ini, metode lbp dan svm digunakan untuk melakukan klasifikasi citra lima jenis kayu yaitu bayur, cempaka, damar, meranti, dan merbau. citra jenis kayu tersebut diambil secara manual menggunakan kamera digital dengan jarak pengambilan 20 cm tanpa penambahan cahaya. untuk menghitung ciri tekstur lbp, perangkat lunak dikembangkan menggunakan visual studio 2015 dengan bahasa program c#. metode ekstraksi ciri tekstur yang digunakan dalam penelitian ini adalah lbp dengan indikator yaitu mean, standar deviasi, skewness, energi, dan entropi. kerangka penelitian tahapan penelitian meliputi pengambilan citra jenis kayu, pra pengolahan, perancangan dan pembuatan aplikasi klasifikasi jenis kayu, ekstraksi ciri tekstur menggunakan lbp, pelatihan dan pengujian menggunakan svm, dan hasil akhir adalah hasil klasifikasi. hasil pemotongan citra kayu bayur, cempaka, damar, meranti, dan merbau dengan ukuran 200 x 200 piksel disajikan dalam gambar 3. hasil pemotongan citra kayu perangkat lunak klasifikasi jenis kayu ini dikembangkan menggunakan visual studio 2015 dengan bahasa programnya adalah c#, dan basis data yang digunakan adalah microsoft acces. sedangkan proses pengujian dilakukan untuk mengklasifikasi atau menentukan jenis kayu sesuai dengan citra aslinya sehingga dapat diketahui prosentase tingkat akurasi benar dan salah terhadap hasil klasifikasi citra. hasil ekstraksi ciri lbp dari masing-masing citra kayu dijadikan sebagai data input untuk melakukan pengujian menggunakan svm sehingga didapatkan nilai akurasi tertinggi dari proses klasifikasi. hasil klasifikasi citra jenis kayu bayur, cempaka, damar, meranti, dan merbau akan diketahui nilai akurasi tertingginya pada nilai parameter sigma tertentu yang digunakan dalam pengujian.hasil klasifikasi dilakukan terhadap citra kayu bayur, cempaka, damar, meranti, dan merbau dengan jumlah masing-masing 150 citra, sehingga total data seluruh adalah 750 citra. hasil pengujian citra kayu bayur, cempaka, damar, meranti, dan merbau dengan jumlah data masing-masing 30 citra menggunakan parameter sigma 0,3 dengan jarak ketetanggaan r=1 disajikan dalam tabel 1.tabel 1. dari hasil pengujian 150 citra kayu yang disajikan dalam tabel 1 di atas, terlihat bahwa 137 citra kayu teridentifikasi benar sesuai dengan gambar aslinya, dan 13 citra teridentifikasi salah tidak sesuai dengan gambar aslinya. tingkat keakuratan klasifikasi citra kayu bayur, cempaka, damar, meranti, dan merbau yang diambil dengan jarak 20 cm berdasarkan pengujian data menggunakan jarak ketetanggaan r=1 adalah sebesar 91,3% terletak pada parametersigma 0,3. nilai error terkecil berdasarkan hasil klasifikasi citra kayu bayur, cempaka, damar, meranti, dan merbau adalah sebesar 8,7% yang berada pada parameter sigma 0,3.

Gambar 4. 8 Ringkasan Sistem *Compression Rate* 30% id-14 dengan *Stemming*

dewasa ini hutan tanaman yang menanam berbagai jenis kayu baik dari jenis unggulan setempat (native species) maupun dari jenis eksotik (exotic species) makin berkembang, contohnya hutan rakyat, hutan kemasyarakatan, dan hutan industri. keterbatasan kemampuan manusia dalam mengidentifikasi jenis kayu yang belum terampil secara visual terkadang berpengaruh terhadap hasil yang diinginkan bagi dunia industri. keuntungan dari metode lbp adalah invarian untuk perubahan skala abu-abu monotonik, kompleksitas komputasi yang rendah dan multi skala yang nyaman. hasil akhir menunjukkan bahwa system dapat mendeteksi adanya cacat pada kayu dengan tingkat akurasi tertinggi adalah 89,4%, far sebesar 7,6% dan frr sebesar 3%, dengan waktu komputasi rata-rata sistem sebesar 0,3069 detik [5]. klasifikasi jenis kayu jati, sengon, mahoni dan mindi pada penelitian terdahulu telah dilakukan menggunakan back-propagation neural network berdasarkan fitur gray level co□occurrence matrices. hasil penelitian ini menunjukkan nilai rata□rata akurasi yaitu 96.13% [6].penelitian mengenai ekstraksi citra kayu juga telah dilakukan sebelumnya pada jenis kayu jati, mahoni, mindi, dan albasia untuk menentukan fitur dan arah sudut yang lebih efisien dan efektif dalam mengidentifikasi spesies kayu. jenis fitur yang diuji dalam penelitian adalah angular second moment (asm), contrast, idm/homogenitas, entropi, korelasi dengan arah komputasi yakni 0, 45, 90, dan 135 derajat glcm. mesin pembelajaran untuk melakukan klasifikasi tekstur citra yang dapat digunakan diantaranya adalah logika fuzzy, jaringan syaraf tiruan, algoritma genetika, dan support vector machine (svm) [8]. penelitian ini menghasilkan tingkat pengenalan terbaik yakni 87,5% berada pada jarak pengambilan 20 cm dengan jarak piksel tetangga d=2 pada arah glcm 135 derajat. sedangkan penambahan kanal warna yakni hue, saturation, dan value (hsv) pada klasifikasi citra daging ini, menghasilkan klasifikasi terbaik yakni 75,6% yang berada pada arah glcm 45 derajat dengan jarak piksel tetangga d=3 dan arah glcm 135 derajat dengan jarak piksel tetangga d=2. dalam penelitian ini, metode lbp dan svm digunakan untuk melakukan klasifikasi citra lima jenis kayu yaitu bayur, cempaka, damar, meranti, dan merbau. citra jenis kayu tersebut diambil secara manual menggunakan kamera digital dengan jarak pengambilan 20 cm tanpa penambahan cahaya. untuk menghitung ciri tekstur lbp, perangkat lunak dikembangkan menggunakan visual studio 2015 dengan bahasa program c#. metode ekstraksi ciri tekstur yang digunakan dalam penelitian ini adalah lbp dengan indikator yaitu mean, standar deviasi, skewness, energi, dan entropi. kerangka penelitian tahapan penelitian meliputi pengambilan citra jenis kayu, pra pengolahan, perancangan dan

pembuatan aplikasi klasifikasi jenis kayu, ekstraksi ciri tekstur menggunakan lbp, pelatihan dan pengujian menggunakan svm, dan hasil akhir adalah hasil klasifikasi. hasil pemotongan citra kayu bayur, cempaka, damar, meranti, dan merbau dengan ukuran 200 x 200 piksel disajikan dalam gambar 3. hasil pemotongan citra kayu perangkat lunak klasifikasi jenis kayu ini dikembangkan menggunakan visual studio 2015 dengan bahasa programnya adalah c#, dan basis data yang digunakan adalah microsoft acces. sedangkan proses pengujian dilakukan untuk mengklasifikasi atau menentukan jenis kayu sesuai dengan citra aslinya sehingga dapat diketahui prosentase tingkat akurasi benar dan salah terhadap hasil klasifikasi citra. hasil ekstraksi ciri lbp dari masing-masing citra kayu dijadikan sebagai data input untuk melakukan pengujian menggunakan svm sehingga didapatkan nilai akurasi tertinggi dari proses klasifikasi. hasil klasifikasi citra jenis kayu bayur, cempaka, damar, meranti, dan merbau akan diketahui nilai akurasi tertingginya pada nilai parameter sigma tertentu yang digunakan dalam pengujian.hasil klasifikasi dilakukan terhadap citra kayu bayur, cempaka, damar, meranti, dan merbau dengan jumlah masing-masing 150 citra, sehingga total data seluruh adalah 750 citra. hasil pengujian citra kayu bayur, cempaka, damar, meranti, dan merbau dengan jumlah data masing-masing 30 citra menggunakan parameter sigma 0,3 dengan jarak ketetanggaan r=1 disajikan dalam tabel 1.tabel 1. dari hasil pengujian 150 citra kayu yang disajikan dalam tabel 1 di atas, terlihat bahwa 137 citra kayu teridentifikasi benar sesuai dengan gambar aslinya, dan 13 citra teridentifikasi salah tidak sesuai dengan gambar aslinya. tingkat keakuratan klasifikasi citra kayu bayur, cempaka, damar, meranti, dan merbau yang diambil dengan jarak 20 cm berdasarkan pengujian data menggunakan jarak ketetanggaan r=1 adalah sebesar 91,3% terletak pada parametersigma 0,3. nilai error terkecil berdasarkan hasil klasifikasi citra kayu bayur, cempaka, damar, meranti, dan merbau adalah sebesar 8,7% yang berada pada parameter sigma 0,3.

Gambar 4. 9 Ringkasan Sistem *Compression Rate* 30%id-14 tanpa *Stemming*

4.2.4 Tingkat Peringkasan 40%

Pada peringkasan dengan tingkat kompresi sebesar 40% tetap menggunakan uji coba parameter dan dengan derajat keanggotaan yang sama seperti peringkasan dengan tingkat kompresi 10%, 20%, dan 30%, dapat dilihat pada Tabel 4.12 dan Tabel 4.13.

Tabel 4. 12 Hasil rata-rata ROUGE *Compression Rate* 40% dengan *Stemming*

Compression Rate	**ROUGE-N**							
	ROUGE-1				**ROUGE-2**			
	Precision	***Recall***	***F1-Score***	***Accuracy***	***Precision***	***Recall***	***F2-Score***	***Accuracy***
40%	0.6325	0.4568	0.5111	0,5447	0.1750	0.1219	0.1382	0,1484

Tabel 4. 13 Hasil rata-rata ROUGE *Compression Rate* 40% tanpa *Stemming*

Compression Rate	**ROUGE-N**							
	ROUGE-1				**ROUGE-2**			
	Precision	***Recall***	***F1-Score***	***Accuracy***	***Precision***	***Recall***	***F2-Score***	***Accuracy***
40%	0.6323	0.4566	0.5109	0,5445	0.1746	0.1216	0.1379	0,1481

Contoh dari hasil ringkasan manual dan ringkasan sistem pada dokumen id-11 yang berjudul "*Analyzing The Effectiveness Of Collaborative Filtering And Content-Based Filtering Methods In Anime Recommendation Systems*" dapat dilihat pada Gambar 4.10, Gambar 4.11, dan Gambar 4.12.

Di era digital saat ini dimana konsumsi konten melalui platform streaming semakin meningkat, kebutuhan akan sistem rekomendasi yang akurat menjadi semakin penting, khususnya di industri animasi. Penelitian ini berfokus pada penerapan sistem rekomendasi yang dapat membantu pemirsa dengan mudah menavigasi banyaknya konten. Dengan membandingkan metode pemfilteran kolaboratif dan pemfilteran berbasis konten, penelitian ini berupaya menemukan pendekatan optimal untuk memberikan rekomendasi anime. Dari hasil pengujian A/B dan analisis lebih lanjut, ditemukan bahwa Collaborative Filtering efektif dalam memberikan rekomendasi berdasarkan kesamaan minat antar pengguna. Di sisi lain, pemfilteran berbasis konten menawarkan keuntungan dalam mempersonalisasi rekomendasi berdasarkan karakteristik konten. Selain itu, mengintegrasikan teknik ini ke dalam aplikasi seluler akan memperkaya pengalaman pengguna, memungkinkan mereka menerima rekomendasi lebih cepat dan interaktif. Penelitian ini berkontribusi pada pengembangan sistem rekomendasi yang lebih intuitif dan responsif, mendorong pertumbuhan industri streaming anime dengan meningkatkan kepuasan dan retensi pengguna.

Gambar 4. 10 Ringkasan Manual id-11

dalam era konten digital yang semakin berkembang, industri hiburan telah melihat peningkatan yang signifikan dalam platform streaming. setiap platform streaming memiliki ribuan konten yang tersedia, termasuk film, acara tv, dan anime. namun dengan berbagai pilihan yang ada, pengguna sering kali menghabiskan waktu berjam-jam untuk mencari konten yang sesuai dengan minat dan preferensi mereka.dalam konteks ini, penting bagi bisnis streaming untuk memahami kebutuhan dan preferensi mereka. melalui penilaian dan ulasan yang diberikan oleh pengguna, bisnis dapat mengidentifikasi konten yang paling disukai dan mengoptimalkan pengalaman penonton. salah satu aspek kunci dalam menciptakan lingkungan streaming yang sukses adalah memberikan rekomendasi yang akurat dan relevan. namun, tantangan muncul ketika pengguna harus menjelajahi ratusan atau bahkan ribuan konten anime yang tersedia. oleh karena itu, sebuah bisnis perlu menghadapi tantangan ini dan menyediakan solusi yang dapat membantu menemukan konten yang sesuai dengan minat mereka.dalam hal ini, implementasi sistem rekomendasi berdasarkan preferensi pengguna menjadi sangat penting. dengan menganalisis data penilaian, preferensi, dan riwayat penonton, bisnis dapat memberikan saran yang personal dan relevan. hal ini tidak hanya memberikan pengalaman pengguna yang lebih baik, tetapi juga dapat meningkatkan waktu yang dihabiskan oleh pengguna di platform streaming dan meningkatkan pendapatan bisnis.dengan memahami preferensi dan kebutuhan pengguna serta memberikan rekomendasi yang tepat, bisnis streaming dapat menciptakan lingkungan yang menguntungkan baik bagi pengguna maupun bisnis itu sendiri.metode penelitian ini terdiri dari beberapa tahap. pada tahap pertama, dilakukan pengumpulan data anime yang diperoleh dari dataset yang tersedia pada situs web kaggle datasetscooperunion. dataset ini berisi informasi tentang judul anime, genre, rating, ulasan, produser, dan atribut lainnya. setelah itu, dilakukan eksplorasi data untuk memahami karakteristik dataset, termasuk distribusi rating anime, jumlah ulasan, genre yang paling umum, dan atribut lainnya. dilanjutkan dengan analisis statistik deskriptif untuk mendapatkan wawasan lebih lanjut tentang data anime, termasuk perhitungan rata-rata, median, sebaran data, dan visualisasi data menggunakan histogram, scatter plot, atau diagram batang. tahap selanjutnya adalah pra-pemrosesan data, yang melibatkan penanganan missing values, penghapusan data yang tidak valid atau tidak relevan, dan konversi variabel kategorikal kedalam bentuk numerik. model collaborative filtering dilatih menggunakan algoritma seperti singular value decomposition (svd) atau alternating least squares (als) untuk menemukan pola dan kesamaan antara pengguna, sementara model content-based filtering mempertimbangkan atribut□atribut anime seperti genre, produser, atau tipe konten. pengguna secara acak dibagi menjadi dua kelompok yakni kelompok yang menerima rekomendasi dari model pemfilteran kolaboratif, dan kelompok lainnya menerima rekomendasi dari model pemfilteran berbasis konten. hasil pengujian a/b dan umpan balik yang diterima memberikan informasi penting tentang kualitas dan efektivitas sistem usulan yang dikembangkan. alur perancangan sistem rekomendasi anime disajikan seperti pada gambar 1.pengumpulan data dalam penelitian ini sumber data utama yang digunakan berasal dari dataset yang tersedia pada platform kaggle tepatnya pada anime recommendations-database. dataset ini dipilih karena memberikan dataset anime yang komprehensif dan detail, merupakan aspek yang sangat penting untuk mendukung analisis dan implementasi sistem rekomendasi yang dibahas dalam penelitian ini. dataset tersebut berisi berbagai informasi tentang anime, antara lain judul, genre, rating, review, produser, dan beberapa atribut terkait lainnya. dengan informasi tersebut, peneliti dapat melakukan analisis mendalam mengenai preferensi dan tren pemirsa anime, serta bagaimana setiap atribut suatu anime dapat mempengaruhi pilihan dan rekomendasi yang akan diberikan kepada pengguna. eksplorasi data sebelum melanjutkan analisis dan pengolahan data, penting untuk memahami struktur dan karakteristik dasar kumpulan data yang kita miliki. distribusi genre anime disajikan pada gambar 6.pada tahap preprocessing

data, langkah pertama yakni menangani missing value dan memastikan tidak ada gap pada dataset. mengingat relevansi dalam sistem pemberi rekomendasi, hanya pengguna yang memberikan minimal 50 peringkat yang dipertimbangkan untuk analisis lebih lanjut. data tersebut kemudian diubah menjadi tabel pivot untuk memfasilitasi pembentukan matriks renggang, yang penting untuk perhitungan kesamaan. selain itu, penulis juga mengklarifikasi nama anime dengan menghapus karakter jepang dan simbol khusus, sehingga meningkatkan kualitas dan keterbacaan data. di era digital saat ini, mobile apps telah menjadi platform yang efektif untuk memberikan solusi dan layanan kepada pengguna. mengikuti tren ini, penerapan teknik pemfilteran kolaboratif dan pemfilteran berbasis konten telah diintegrasikan ke dalam aplikasi seluler. dengan menggunakan aplikasi seluler sebagai media, proses rekomendasi menjadi lebih cepat dan interaktif, memberikan pengguna akses instan ke konten yang paling mereka minati. untuk memahami efektivitas antara pemfilteran kolaboratif (cf) dan pemfilteran berbasis konten (cbf) dalam konteks sistem rekomendasi anime, penulis melakukan pengujian a/b. hasil survei kemudian dianalisis menggunakan metode statistik, seperti uji-t atau anova, untuk mengetahui apakah terdapat perbedaan respon yang signifikan antara kedua kelompok. eksperimen ini penting untuk memandu keputusan dalam memilih metode rekomendasi yang paling tepat, dengan mempertimbangkan konteks pengguna dan tujuan yang diinginkan. dalam penelitian ini, peneliti menganalisis dan membandingkan dua metode yang umum digunakan dalam sistem rekomendasi anime, yaitu collaborative filtering (cf) dan content based filtering (cbf), melalui eksperimen a/b. hasilnya menunjukkan bahwa meskipun cf efektif dalam membuat rekomendasi berdasarkan preferensi serupa pengguna, cbf tampaknya memberikan kepuasan yang lebih tinggi dengan berfokus pada karakteristik konten animasi untuk membuat rekomendasi yang dipersonalisasi.

Gambar 4. 11 Ringkasan Sistem *Compression Rate* 40% id-11 dengan *Stemming*

dalam era konten digital yang semakin berkembang, industri hiburan telah melihat peningkatan yang signifikan dalam platform streaming. setiap platform streaming memiliki ribuan konten yang tersedia, termasuk film, acara tv, dan anime. namun dengan berbagai pilihan yang ada, pengguna sering kali menghabiskan waktu berjam-jam untuk mencari konten yang sesuai dengan minat dan preferensi mereka.dalam konteks ini, penting bagi bisnis streaming untuk memahami kebutuhan dan preferensi mereka. melalui penilaian dan ulasan yang diberikan oleh pengguna, bisnis dapat mengidentifikasi konten yang paling disukai dan mengoptimalkan pengalaman penonton. salah satu aspek kunci dalam menciptakan lingkungan streaming yang sukses adalah memberikan rekomendasi yang akurat dan relevan. namun, tantangan muncul ketika pengguna harus menjelajahi ratusan atau bahkan ribuan konten anime yang tersedia. oleh karena itu, sebuah bisnis perlu menghadapi tantangan ini dan menyediakan solusi yang dapat membantu menemukan konten yang sesuai dengan minat mereka.dalam hal ini, implementasi sistem rekomendasi berdasarkan preferensi pengguna menjadi sangat penting. dengan menganalisis data penilaian, preferensi, dan riwayat penonton, bisnis dapat memberikan saran yang personal dan relevan. hal ini tidak hanya memberikan pengalaman pengguna yang lebih baik, tetapi juga dapat meningkatkan waktu yang dihabiskan oleh pengguna di platform streaming dan meningkatkan pendapatan bisnis.dengan memahami preferensi dan kebutuhan pengguna serta memberikan rekomendasi yang tepat, bisnis streaming dapat menciptakan lingkungan yang menguntungkan baik bagi pengguna maupun bisnis itu sendiri.metode penelitian ini terdiri dari beberapa tahap. pada tahap pertama, dilakukan pengumpulan data anime yang diperoleh dari dataset yang tersedia pada situs web kaggle datasetscooperunion. dataset ini berisi informasi tentang judul anime, genre, rating, ulasan, produser, dan atribut lainnya. setelah itu, dilakukan eksplorasi data untuk memahami karakteristik dataset, termasuk distribusi rating anime, jumlah ulasan, genre yang paling umum, dan atribut lainnya. dilanjutkan dengan analisis statistik deskriptif untuk mendapatkan wawasan lebih lanjut tentang data anime, termasuk perhitungan rata-rata, median, sebaran data, dan visualisasi data menggunakan histogram, scatter plot, atau diagram batang. tahap selanjutnya adalah pra-pemrosesan data, yang melibatkan penanganan missing values, penghapusan data yang tidak valid atau tidak relevan, dan konversi variabel kategorikal kedalam bentuk numerik. model collaborative filtering dilatih menggunakan algoritma seperti singular value decomposition (svd) atau alternating least squares (als) untuk menemukan pola dan kesamaan antara pengguna, sementara model content-based

filtering mempertimbangkan atribut□atribut anime seperti genre, produser, atau tipe konten. pengguna secara acak dibagi menjadi dua kelompok yakni kelompok yang menerima rekomendasi dari model pemfilteran kolaboratif, dan kelompok lainnya menerima rekomendasi dari model pemfilteran berbasis konten. hasil pengujian a/b dan umpan balik yang diterima memberikan informasi penting tentang kualitas dan efektivitas sistem usulan yang dikembangkan. alur perancangan sistem rekomendasi anime disajikan seperti pada gambar 1.pengumpulan data dalam penelitian ini sumber data utama yang digunakan berasal dari dataset yang tersedia pada platform kaggle tepatnya pada anime recommendations-database. dataset ini dipilih karena memberikan dataset anime yang komprehensif dan detail, merupakan aspek yang sangat penting untuk mendukung analisis dan implementasi sistem rekomendasi yang dibahas dalam penelitian ini. dataset tersebut berisi berbagai informasi tentang anime, antara lain judul, genre, rating, review, produser, dan beberapa atribut terkait lainnya. dengan informasi tersebut, peneliti dapat melakukan analisis mendalam mengenai preferensi dan tren pemirsa anime, serta bagaimana setiap atribut suatu anime dapat mempengaruhi pilihan dan rekomendasi yang akan diberikan kepada pengguna. eksplorasi data sebelum melanjutkan analisis dan pengolahan data, penting untuk memahami struktur dan karakteristik dasar kumpulan data yang kita miliki. distribusi genre anime disajikan pada gambar 6.pada tahap preprocessing data, langkah pertama yakni menangani missing value dan memastikan tidak ada gap pada dataset. mengingat relevansi dalam sistem pemberi rekomendasi, hanya pengguna yang memberikan minimal 50 peringkat yang dipertimbangkan untuk analisis lebih lanjut. data tersebut kemudian diubah menjadi tabel pivot untuk memfasilitasi pembentukan matriks renggang, yang penting untuk perhitungan kesamaan. selain itu, penulis juga mengklarifikasi nama anime dengan menghapus karakter jepang dan simbol khusus, sehingga meningkatkan kualitas dan keterbacaan data. mengikuti tren ini, penerapan teknik pemfilteran kolaboratif dan pemfilteran berbasis konten telah diintegrasikan ke dalam aplikasi seluler. dengan menggunakan aplikasi seluler sebagai media, proses rekomendasi menjadi lebih cepat dan interaktif, memberikan pengguna akses instan ke konten yang paling mereka minati. untuk memahami efektivitas antara pemfilteran kolaboratif (cf) dan pemfilteran berbasis konten (cbf) dalam konteks sistem rekomendasi anime, penulis melakukan pengujian a/b. setelah menerima daftar rekomendasi, pengguna diminta menilai kepuasan mereka terhadap rekomendasi yang diterima melalui survei skala likert. hasil survei kemudian dianalisis menggunakan metode statistik, seperti uji-t atau anova, untuk mengetahui apakah terdapat perbedaan respon yang signifikan antara kedua kelompok. eksperimen ini penting untuk memandu keputusan dalam memilih metode rekomendasi yang paling tepat, dengan mempertimbangkan konteks pengguna dan tujuan yang diinginkan. dalam penelitian ini, peneliti menganalisis dan membandingkan dua metode yang umum digunakan dalam sistem rekomendasi anime, yaitu collaborative filtering (cf) dan content based filtering (cbf), melalui eksperimen a/b. hasilnya menunjukkan bahwa meskipun cf efektif dalam membuat rekomendasi berdasarkan preferensi serupa pengguna, cbf tampaknya memberikan kepuasan yang lebih tinggi dengan berfokus pada karakteristik konten animasi untuk membuat rekomendasi yang dipersonalisasi.

Gambar 4. 12 Ringkasan Sistem *Compression Rate* 40% id-11 tanpa *Stemming*

4.2.5 Tingkat Peringkasan 50%

Pada peringkasan dengan tingkat kompresi sebesar 50% tetap sama dengan menggunakan parameter yang sama seperti pada peringkasan dengan tingkat kompresi sebelumnya pada 10%, 20%, 30%, dan 40%. Nilai rata-rata dari hasil evaluasi matriks ROUGE-N pada peringkasan dengan tingkat kompresi 50% dapat dilihat pada Tabel 4.14 dan Tabel 4.15.

Tabel 4. 14 Hasil rata-rata ROUGE *Compression Rate* 50% dengan *Stemming*

Compression Rate	**ROUGE-N**							
	ROUGE-1				**ROUGE-2**			
	Precision	***Recall***	***F1-Score***	***Accuracy***	***Precision***	***Recall***	***F2-Score***	***Accuracy***
50%	0.7073	0.4121	0.5034	0,5597	0.2198	0.1228	0.1520	0,1713

Tabel 4. 15 Hasil rata-rata ROUGE *Compression Rate* 50% tanpa *Stemming*

Compression Rate	**ROUGE-N**							
	ROUGE-1				**ROUGE-2**			
	Precision	***Recall***	***F1-Score***	***Accuracy***	***Precision***	***Recall***	***F2-Score***	***Accuracy***
50%	0.7069	0.4118	0.5031	0,5594	0.2192	0.1225	0.1516	0,1709

Contoh dari hasil ringkasan manual dan ringkasan sistem pada dokumen id-39 yang berjudul “Analisis Sentimen Respon Masyarakat Terhadap Kabar Harian Covid-19 Pada Twitter Kementerian Kesehatan Dengan Metode Klasifikasi Naive Bayes” dapat dilihat pada Gambar 4.13, Gambar 4.14 dan Gambar 15.

Kementerian Kesehatan Republik Indonesia berperan sebagai gugus terdepan dalam penanganan covid-19 di Indonesia selalu menyajikan kabar harian untuk menyediakan informasi mengenai kasus pandemi covid-19 perharinya melalui akun twitter milik Kementerian Kesehatan Republik Indonesia. Namun, tidak semua tweet harian yang diberikan oleh akun twitter Kementerian Kesehatan Republik Indonesia dapat dikonsumsi dan direspon masyarakat dengan selaras. Adapun masalah ini dapat diatasi dengan melakukan penelitian di bidang Analisis Sentiment, yang mana merupakan bidang penelitian yang berfokus kepada studi komputasi atas opini, tingkah laku, dan emosi terhadap suatu entitas yang dituangkan dalam bentuk teks. Penelitian ini dilakukan untuk mengetahui bagaimana hasil analisis sentimen terkait respon masyarakat dari kabar harian Covid_x0002_19 dari twitter Kementerian Kesehatan Republik Indonesia dan mengklassifikasikannya menjadi tiga kelas yaitu positif, negatif, dan netral menggunakan metode klasifikasi Naïve Bayes Classifier, sehingga penelitian ini diharapkan dapat menghasilkan sentimen dengan kelas yang cenderung positif, negatif, atau netral. Hasil dari kesimpulan penelitian ini nantinya dapat dilihat dalam bentuk grafik. Penelitian ini juga melakukan pengujian akurasi, pengujian precision, dan recall, f-1 score untuk memastikan ke akuratan dari penelitian.

Gambar 4. 13 Ringkasan Manual id-39

informasi yang sedang menjadi perbincangan utama akhir- akhir ini berasal dari masyarakat dunia yang sedang di gemparkan dengan adanya pandemi covid-19 (corona virus disease 2019) yang tengah berlangsung sejak awal maret lalu hingga saat ini, pandemi ini tidak hanya mewabah di indonesia tetapi juga menyebar hampir di seluruh penjuru dunia. belum tuntasnya pelaksanaan pemutusan rantai penyebaran virus covid-19 dari pemerintah hingga saat ini, tentu membuat kalangan masyarakat menjadi resah, was-was hingga ketakutan dengan adanya pandemi ini. seiring dengan terus berlanjutnya pandemi covid-19 hingga saat ini, beragam cara telah dilakukan oleh pemerintah hingga kalangan masyarakat demi memutus rantai penyebaran virus covid- 19. kementerian kesehatan republik indonesia berperan sebagai gugus terdepan dalam penanganan covid-19 di indonesia selalu menyajikan kabar harian untuk menyediakan informasi mengenai kasus pandemi covid-19 perharinya melalui akun twitter milik kementerian kesehatan republik indonesia [9]. namun, tidak semua tweet harian yang diberikan oleh twitter kementerian kesehatan republik indonesia dapat dikonsumsi dan direspon masyarakat dengan selaras. akun twitter milik kementerian kesehatan republik indonesia dibanjiri dengan beragam komentar yang muncul dari kalangan masyarakat pengguna twitter, komentar yang diberikan oleh pengguna twitter ini merupakan data teks yang dapat digali kembali, diolah, dan dimanfaatkan dan dijadikan sebagai bahan penelitian untuk berbagai keperluan di bidang pengembangan informasi. penelitian ini

menggunakan data teks berbahasa indonesia dari akun twitter kementerian kesehatan republik indonesia dengan metode klasifikasi na ve bayes classifier. berdasarkan pemaparan masalah tersebut, penelitian ini dilaksanakan untuk mengetahui bagaimana hasil analisis sentimen terkait respon masyarakat dari kabar harian covid-19 dari twitter kementerian kesehatan republik indonesia dan mengklassifikasikannya menjadi tiga kelas yaitu positif, negatif, dan netral menggunakan metode klasifikasi na ve bayes classifier, sehingga penelitian ini diharapkan dapat menghasilkan sentimen dengan kelas yang cenderung positif, negatif, atau netral. media sosial twitter merupakan satu dari beberapa bagian perkembangan media komunikasi yang diciptakan dengan tujuan agar pengguna bisa memberikan ekspresi, pendapat, aspirasi, kritik, serta bertukar informasi mengenai suatu informasi yang sedang menjadi perbincangan utama, tanpa keterbatasan waktu dan ruang. hal tersebut menjadikan twitter sebagai salah satu sumber data text yang dapat digali dan dimanfaatkan untuk berbagai keperluan penelitian di bidang teknologi informasi. twitter telah menyiapkan structured text (xml dan daftar kategori) di dalamnya selalu terdapat informasi username, timestamp, text, retweet, favorite dan informasi lainnya, hal ini dapat dilihat dari proses implementasi metadata di setiap tweet. analisis sentimen atau opinion mining merupakan salah satu bagian dari bidang studi text mining yang disebut sebagai studi komputasi mengenai pendapat rakyat, sentimen, penilaian, sikap, dan emosi terhadap entitas dan atribut yang fokusnya adalah mengeksraksi, mengidentifikasi dan meneliti dan menemukan karakteristik sentimen atau kecenderungan seseorang terhadap sebuah opini dalam sebuah masalah atau objek oleh seseorang dalam unit teks menggunakan metode nlp (natural language processing), statistik atau machine learning.naive bayes classifier digunakan untuk mencari nilai probabilitas tertinggi dalam teknik klasifikasi data uji pada kategori yang paling tepat. saat proses klasifikasi algoritma naive bayes akan mencari probabilitas tertinggi dari semua kategori dokumen yang diujikan (vmap). dataset dikumpulkan melalui pengambilan data dari twitter kementrian kesehatan republik indonesia menggunakan python dan memanfaatkan fasilitas twitterscrapper, berupa komentar tweet berbahasa indonesia sebanyak 2397 data yang diambili mulai dari 1 maret 2020 hingga 30 oktober 2020 dengan kata kunci covid-19 dan kemenkes dalam bentuk file csv.preprocessing merupakan tahap awal pada proses pengolahan data teks hasil ekstraksi crawling sebelum diolah lebih lanjut ke tahap proses klasifikasi hingga nanti masuk ketahap uji dan evaluasi. pengujian ini dilakukan untuk bertujuan memastikan bahwa penelitian yang dilaksanakan dengan metode naive bayes classifier menghasilkan nilai akurasi dengan keakuratan yang baik. kelas negatif melakukan peritungan confussion matrix menggunakan library sckit.learn yang merupakan library yang sama dengan proses klasifikasi menggunakan naive bayes classifier. hasil dari model training diklasifikasikan dengan testing mengasilkan matrix dengan ukuran 3x3 sebagai representatif kelas aktual dan kelas prediksi.tabel 1 confussion matrix. merupakan hasil dari prediksi menggunakan mesin klasifikasi naive bayes classifier yang akan diukur performa dari tiap-tiap kelas dengan mengitung precision, recall, dan f1-score. berdasarkan tabel didapatkan hasil confusiion matrix dari mesin melaui library sckit.learn, dapat terlihat jumlah kelas kata yang diprediksi oleh mesin yaitu: sebanyak 24 kelas kata terprediksi benar bernilai positif (tp), 3 kelas kata positif terprediksi netral (pn), 16 kelas kata positif terprediksi negatif (fp), 7 kelas kata netral terprediksi sebagai kelas kata positif (np), 4 kelas kata terprediksi netral (tn), 16 kelas kata netral terprediksi negatif (nng), 33 kelas kata negatif terprediksi positif (fn), 30 kelas kata negatif terprediksi netral (ngn), dan 347 kelas kata negatif terprediksi negatif (tng). persentase ini dibuat untuk memudahkan pembaca serta mengukur berapa besar sentiment yang diberikan oleh masyarakat terhadap respon kabar harian covid-19 pada twitter kementrian kesehatan ri berdasarkan proses labeling. pada gambar ini menunjukkan bahwa hasil dari sentimen analisis yang telah dilakukan menghasilkan persentase 85% negatif, 4% netral dan 11% positif. dari penelitian yang telah

dilaksanakan dengan penggunaan metode machine learning yaitu algoritma naive bayes clasifier pada dataset kabar harian covid-19 pada twitter kementrian kesehatan republik indonesia dengan kata kunci kemenkes dan covid-19 telah didapatkan dataset hasil crawling sebanyak 2397 dataset, yang kemudian dilakukan proses preprocessing kemudian dataset tersebut diolah untuk proses selanjutnya untuk mendapatkan kelas sentimen. selanjutnya setelah melakukan klasifikaasi sentimen, didapatkan hasil klassifikasi sentimen dengan tiga kelas, yaitu kelas positif sebanyak 11%, kelas negatif sebanyak 85%, dan kelas netral sebanyak 4%. data hasil klasifikasi ini diperoleh dengan membagi data menjadi data training dan teting, dengan ketentuan jumlah data training sebanyak 80% dan data testing sebanyak 20%. dari hasil klasifikasi na ve bayes classifier dan pengujian akurasi, precession, recall, dan f1-score, dengan demikian dapat disimpulkan bahwa penelitian ini menghasilkan sentimen masyarakat pengguna twitter mengenai respon masyarakat mengenai kabar harian covid-19 yang diberikan oleh twitter kementrian kesehatan republik indonesia dengan presentase kelas sentimen negatif sebesaar 77%.

Gambar 4. 14 Ringkasan Sistem *Compression Rate* 50%id-39 dengan *Stemming*

informasi yang sedang menjadi perbincangan utama akhir- akhir ini berasal dari masyarakat dunia yang sedang di gemparkan dengan adanya pandemi covid-19 (corona virus disease 2019) yang tengah berlangsung sejak awal maret lalu hingga saat ini, pandemi ini tidak hanya mewabah di indonesia tetapi juga menyebar hampir di seluruh penjuru dunia. belum tuntasnya pelaksanaan pemutusan rantai penyebaran virus covid-19 dari pemerintah hingga saat ini, tentu membuat kalangan masyarakat menjadi resah, was-was hingga ketakutan dengan adanya pandemi ini. seiring dengan terus berlanjutnya pandemi covid-19 hingga saat ini, beragam cara telah dilakukan oleh pemerintah hingga kalangan masyarakat demi memutus rantai penyebaran virus covid- 19. kementerian kesehatan republik indonesia berperan sebagai gugus terdepan dalam penanganan covid-19 di indonesia selalu menyajikan kabar harian untuk menyediakan informasi mengenai kasus pandemi covid-19 perharinya melalui akun twitter milik kementerian kesehatan republik indonesia [9]. namun, tidak semua tweet harian yang diberikan oleh twitter kementerian kesehatan republik indonesia dapat dikonsumsi dan direspon masyarakat dengan selaras. akun twitter milik kementerian kesehatan republik indonesia dibanjiri dengan beragam komentar yang muncul dari kalangan masyarakat pengguna twitter, komentar yang diberikan oleh pengguna twitter ini merupakan data teks yang dapat digali kembali, diolah, dan dimanfaatkan dan dijadikan sebagai bahan penelitian untuk berbagai keperluan di bidang pengembangan informasi. penelitian ini menggunakan data teks berbahasa indonesia dari akun twitter kementerian kesehatan republik indonesia dengan metode klasifikasi na ve bayes classifier. berdasarkan pemaparan masalah tersebut, penelitian ini dilaksanakan untuk mengetahui bagaimana hasil analisis sentimen terkait respon masyarakat dari kabar harian covid-19 dari twitter kementerian kesehatan republik indonesia dan mengklassifikasikannya menjadi tiga kelas yaitu positif, negatif, dan netral menggunakan metode klasifikasi na ve bayes classifier, sehingga penelitian ini diharapkan dapat menghasilkan sentimen dengan kelas yang cenderung positif, negatif, atau netral. media sosial twitter merupakan satu dari beberapa bagian perkembangan media komunikasi yang diciptakan dengan tujuan agar pengguna bisa memberikan ekspresi, pendapat, aspirasi, kritik, serta bertukar informasi mengenai suatu informasi yang sedang menjadi perbincangan utama, tanpa keterbatasan waktu dan ruang. hal tersebut menjadikan twitter sebagai salah satu sumber data text yang dapat digali dan dimanfaatkan untuk berbagai keperluan penelitian di bidang teknologi informasi. twitter telah menyiapkan structured text (xml dan daftar kategori) di dalamnya selalu terdapat informasi username, timestamp, text, retweet, favorite dan informasi lainnya, hal ini

dapat dilihat dari proses implementasi metadata di setiap tweet. analisis sentimen atau opinion mining merupakan salah satu bagian dari bidang studi text mining yang disebut sebagai studi komputasi mengenai pendapat rakyat, sentimen, penilaian, sikap, dan emosi terhadap entitas dan atribut yang fokusnya adalah mengeksraksi, mengidentifikasi dan meneliti dan menemukan karakteristik sentimen atau kecenderungan seseorang terhadap sebuah opini dalam sebuah masalah atau objek oleh seseorang dalam unit teks menggunakan metode nlp (natural language processing), statistik atau machine learning.naive bayes classifier digunakan untuk mencari nilai probabilitas tertinggi dalam teknik klasifikasi data uji pada kategori yang paling tepat. saat proses klasifikasi algoritma naive bayes akan mencari probabilitas tertinggi dari semua kategori dokumen yang diujikan (vmap). dataset dikumpulkan melalui pengambilan data dari twitter kementrian kesehatan republik indonesia menggunakan python dan memanfaatkan fasilitas twitterscrapper, berupa komentar tweet berbahasa indonesia sebanyak 2397 data yang diambili mulai dari 1 maret 2020 hingga 30 oktober 2020 dengan kata kunci covid-19 dan kemenkes dalam bentuk file csv.preprocessing merupakan tahap awal pada proses pengolahan data teks hasil ekstraksi crawling sebelum diolah lebih lanjut ke tahap proses klasifikasi hingga nanti masuk ketahap uji dan evaluasi. pengujian ini dilakukan untuk bertujuan memastikan bahwa penelitian yang dilaksanakan dengan metode naive bayes classifier menghasilkan nilai akurasi dengan keakuratan yang baik. kelas negatif melakukan peritungan confussion matrix menggunakan library sckit.learn yang merupakan library yang sama dengan proses klasifikasi menggunakan naive bayes classifier. hasil dari model training diklasifikasikan dengan testing mengasilkan matrix dengan ukuran 3x3 sebagai representatif kelas aktual dan kelas prediksi.tabel 1 confussion matrix. merupakan hasil dari prediksi menggunakan mesin klasifikasi naive bayes classifier yang akan diukur performa dari tiap-tiap kelas dengan mengitung precision, recall, dan f1-score. berdasarkan tabel didapatkan hasil confusiion matrix dari mesin melaui library sckit.learn, dapat terlihat jumlah kelas kata yang diprediksi oleh mesin yaitu: sebanyak 24 kelas kata terprediksi benar bernilai positif (tp), 3 kelas kata positif terprediksi netral (pn), 16 kelas kata positif terprediksi negatif (fp), 7 kelas kata netral terprediksi sebagai kelas kata positif (np), 4 kelas kata terprediksi netral (tn), 16 kelas kata netral terprediksi negatif (nng), 33 kelas kata negatif terprediksi positif (fn), 30 kelas kata negatif terprediksi netral (ngn), dan 347 kelas kata negatif terprediksi negatif (tng). persentase ini dibuat untuk memudahkan pembaca serta mengukur berapa besar sentiment yang diberikan oleh masyarakat terhadap respon kabar harian covid-19

pada twitter kementrian kesehatan ri berdasarkan proses labeling. pada gambar ini menunjukkan bahwa hasil dari sentimen analisis yang telah dilakukan menghasilkan persentase 85% negatif, 4% netral dan 11% positif. dari penelitian yang telah dilaksanakan dengan penggunaan metode machine learning yaitu algoritma naive bayes clasifier pada dataset kabar harian covid-19 pada twitter kementrian kesehatan republik indonesia dengan kata kunci kemenkes dan covid-19 telah didapatkan dataset hasil crawling sebanyak 2397 dataset, yang kemudian dilakukan proses preprocessing kemudian dataset tersebut diolah untuk proses selanjutnya untuk mendapatkan kelas sentimen. selanjutnya setelah melakukan klasifikaasi sentimen, didapatkan hasil klassifikasi sentimen dengan tiga kelas, yaitu kelas positif sebanyak 11%, kelas negatif sebanyak 85%, dan kelas netral sebanyak 4%. data hasil klasifikasi ini diperoleh dengan membagi data menjadi data training dan teting, dengan ketentuan jumlah data training sebanyak 80% dan data testing sebanyak 20%. dari hasil klasifikasi na ve bayes classifier dan pengujian akurasi, precession, recall, dan f1-score, dengan demikian dapat disimpulkan bahwa penelitian ini menghasilkan sentimen masyarakat pengguna twitter mengenai respon masyarakat mengenai kabar harian covid-19 yang diberikan oleh twitter kementrian kesehatan republik indonesia dengan presentase kelas sentimen negatif sebesaar 77%.

Gambar 4. 15 Ringkasan Sistem *Compression Rate* 50%id-39 tanpa *Stemming*

4.3 Pembahasan

Nilai *precision* yang tinggi berarti bahwa ringkasan sistem mencakup informasi yang relevan jika dibandingkan dengan ringkasan manual, artinya semakin tinggi nilai *precision,* semakin sedikit kesalahan yang dibuat oleh sistem dalam memasukkan informasi yang tidak relevan. Nilai *recall* mengukur seberapa banyak informasi penting dari ringkasan manual yang berhasil ditangkap oleh ringkasan sistem. Perbandingan antara *precision* dan *recall* akan menghasilkan nilai *f-score,* mengukur keseimbangan antara relevansi *precison* dan *recall.*

Hasil keseluruhan dari peringkasan teks dengan tingkat kompresi 10%, 20%, 30%, 40% dan 50% dapat dilihat pada Tabel 4.16 dengan *Stemming* dan Tabel 4.17 tanpa *Stemming* dibawah ini.

Tabel 4. 16 Rata-rata hasil evaluasi ROUGE1 dan ROUGE2 dengan *Stemming*

Compresion Rate	**ROUGE-N**							
	ROUGE-1				**ROUGE-2**			
	Precision	***Recall***	***F1-Score***	***Accuracy***	***Precision***	***Recall***	***F2-Score***	***Accuracy***
10%	0.2321	0.6278	0.3271	0,4300	0.0539	0.1414	0.0751	0,0976
20%	0.4053	0.5643	0.4530	0,4848	0.0984	0.1325	0.1080	0,1154
30%	0.5331	0.5034	0.4975	0,5183	0.1369	0.1243	0.1249	0,1306
40%	0.6325	0.4568	0.5111	0,5447	0.1750	0.1219	0.1382	0,1484
50%	0.7073	0.4121	0.5034	0,5597	0.2198	0.1228	0.1520	0,1713

Tabel 4. 17 Rata-rata hasil evaluasi ROUGE1 dan ROUGE2 tanpa *Stemming*

Compresion Rate	**ROUGE-N**							
	ROUGE-1				**ROUGE-2**			
	Precision	***Recall***	***F1-Score***	***Accuracy***	***Precision***	***Recall***	***F2-Score***	***Accuracy***
10%	0.2321	0.6278	0.3271	0,4300	0.0539	0.1414	0.0751	0,0976
20%	0.4049	0.5640	0.4526	0,4844	0.0984	0.1325	0.1080	0,1153
30%	0.5330	0.5033	0.4974	0,5181	01368	01242	0.1248	0,1305
40%	0.6323	0.4566	0.5109	0,5445	0.1746	0.1216	0.1379	0,1481
50%	0.7069	0.4118	0.5031	0,5594	0.2192	0.1225	0.1516	0,1709

Tabel 4.16 dan Tabel 4.17 menunjukkan hasil rata-rata evaluasi keseluruhan menggunakan metriks ROUGE pada berbagai tingkat kompresi teks mulai dari 10% hingga 50% yang menggunakan *Stemming* pada *Preprocessing* dan tanpa menggunakan *Stemming* pada *Preprocessing*nya. Dari hasil uji coba yang diperoleh, terlihat bahwa nilai *precision, recall* dan *f1-score* untuk ROUGE-1, serta *precision, recall,* dan *f2-score* untuk ROUGE-2 mengalami peningkatan seiring bertambahnya tingkat kompresi. Kedua tabel menunjukkan skor yang seimbang

ketika berada pada tingkat kompresi 30%, terlihat bahwa antara nilai *precision* dan *recall* seimbang sebesar 0.5331 dan 0.5034, rata-rata nilai *precision* yang dihasilkan menandakan bahwa lebih dari setengah kata yang muncul dalam ringkasan yang dihasilkan oleh sistem sesuai dengan kata dalam ringkasan referensi. Nilai *recall* yang sedikit lebih rendah, berarti bahwa lebih dari 50% sistem dapat menangkap kata-kata yang relevan dari teks asli dalam ringkasan yang dihasilkan. Nilai *f-score* menunjukkan keseimbangan antara nilai *precision* dan *recall*, yang berarti bahwa sistem cukup baik dalam mempertahankan informasi yang relevan. Akurasi yang dihasilkan sebesar 0.5183 menandakan bahwa model cukup andal dalam menghasilkam ringkasan yang relevan.

F1-Score dan *f2-Score* memiliki perbedaan yang sangat kecil, yang berarti bahwa *stemming* tidak memberikan perubahan yang signifikam terhadap keseimbangan antara *recall* dan *precision,* hanya memberikan pengaruh kurang lebih 0.3%.

Pada penelitian yang pernah dilakukan (Adelia et al., 2019) membuat sistem peringkasan teks menggunakan metode *Bidirectional Gate Recurrent Unit* (BIGRU) dengan dataset 500 jurnal berbahasa indonesia dan abstrak sebagai hasil target ringkasan atau sebagai ringkasan referensinya, menghasilkan rata-rata ROUGE-1 dan ROUGE-2 terbaik sebesar 0.1197 dan 0.0119.

Dari percobaan yang dilakukan diatas bahwa sistem mampu meringkas otomatis dengan peringkasan yang cukup relevan pada tingkat kompresi 30% dengan menggunakan *stemming* pada ROUGE-1. Hasil peringkasan menggunakan *Fuzzy C-Means* dan *Vector Space Model* juga lebih baik dari penelitian yang

menggunakan metode *Bidirectional Gate Recurrent Unit* (BIGRU) dengan dataset dan target ringkasan manual sama.

4.4 Integrasi Islam

Sistem peringkasan teks otomatis yang dibuat menghasilkan ringkasan yang cukup baik dan ringkas pada tingkat kompresi 30% teks nya ringkas dan relevan dengan ringkasan manualnya, dengan hasil tersebut maka pembaca akan lebih paham dengan ringkasan yang dihasilkan karena mengandung cukup banyak yang relevan dengan ringkasan manual, dengan hasil tersebut sesuai dengan hadist yang diriwayatkan oleh Tirmidzi- diriwayatkan oleh Nasa'i- diriwayatkan oleh Abu Daud- diriwayatkan oleh Ahmad bahwa perkataan Rasulullah sallallahu 'alaihi wa sallam adalah jelas (rinci).

عن عائشة رضي الله عنها قالت: كَانَ كَلاَمُ رَسُولِ الله –صلَّى الله عليه وسلَّم– كَلاَمًا فَصلاً يَفْهَمُهُ كُلُّ مَنْ يَسْمَعُهُ [رواه أبو داود واللفظ له، والتُرمذي والنسائي وأحمد].

Dari Aisyah radiyallahu 'anha, ia berkata, "Perkataan Rasulullah sallallahu 'alaihi wa sallam adalah perkataan yang jelas (rinci) dapat dipahami oleh setiap orang yang mendengarnya." (HR. Abu Daud, Tirmidzi, Nasa'i, dan Ahmad)

Perkataan Rasulullah SAW. Yang rinci dan jelas artinya bahwa tidak ada perkataan beliau yang hurufnya bertumpuk satu sama lain dan tidak ada kata-kata yang bercampur dengan kata-kata yang lainnya. Perkataanya jelas dan terang bagi setiap orang yang mendengarnya tanpa bertele-tele, sehingga jika ada orang yang menghitungnya pasti dia akan mampu menghitungnya karena begitu lambatnya pembicaraan Rasulullah SAW. Hal ini disebabkan karena Rasulullah SAW. diberi "Jawami' Al Kalim" dan perkataan dibuat ringkas baginya. Jawami' Al-Kalim

artinya kalimat pendek namun mengandung makna yang komprehensif . Seperti hasil ringkasan oleh sistem ini yang walaupun pendek tetapi mengandung makna yang mudah dipahami.

Alat yang digunakan untuk meringkas teks menjadi sebuah sarana untuk membantu dalam mencapai pemahaman tentang isi dari sebuah artikel, hal ini sesuai dengan sabda Rasulullah sallallahu ‘alaihi wa sallam,

أحبُّ الناسِ إلى اللهِ أَنْفَعُهُمْ لِلنَّاسِ

“Manusia yang paling dicintai Allah adalah yang paling bermanfaat untuk manusia.”(HR At- Tabrani)

Sebaik baiknya manusia adalah yang paling bermanfaat bagi manusia lainnya, sudah sebaiknya kita sebagai manusia saling membantu dan berguna bagi satu sama lain (Febriani, 2024). Hukum asal membantu yaitu mubah, bisa menjadi sunnah jika sesuatu yang bermanfaat itu bisa mengantarkan seseorang lebih dekat dengan Allah SWT. seperti pada hadist tersebut bahwa dalam sistem peringkasan teks yang akan dibuat, dapat dimanfaatkan manusia untuk sarana mencari ilmu dan akan mendapat pahala melalui ilmu yang didapatkannya. Sheikh Yusuf Al-Qaradawi, dalam fatwa terkait teknologi dan informasi, menegaskan bahwa semua inovasi teknologi adalah mubah. Sistem peringkasan yang akan dibuat dapat dianggap sebagai alat yang mubah, karena membantu untuk mempermudah mencari informasi inti dari sebuah artikel jurnal ilmiah. Fatwa lain dari Dar al-Ifta Mesir menekankan bahwa teknologi digital, termasuk sistem peringkasan teks yang akan dibuat, bisa diterima selama

digunakan untuk tujuan yang bermanfaat dan tidak melanggar prinsip-prinsip syariah.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Hasil pengujian yang sudah dilakukan pada sistem peringkasan teks artikel jurnal ilmiah berbahasa indonesia menggunakan *Fuzzy C-Means* dan *Vector Space Model* dengan 100 dokumen artikel jurnal ilmiah menghasilkan nilai rata-rata hasil evaluasi ROUGE-1 dan ROUGE-2, yang ditunjukkan pada Tabel 4.16 dan Tabel 4.17 menggunakan *stemming* dan tanpa *stemming* menghasilkan bahwa sistem memberikan hasil yang relevan untuk menangkap kata-kata serta mengklaster dengan baik, mana kalimat yang cocok dijadikan untuk ringkasan atau kurang cocok berdasarkan hasil perhitungan skor yang dilakukan. Hal ini dibuktikan dengan hasil relevansi rata-rata nilai ROUGE-1 pada tingkat kompresi 30% dengan nilai rata-rata *precision* 0.5331, *recall* 0.5034, *f1-score* 0.4975 dan *accuracy* 0.5183.

Sistem dengan *stemming* menghasilkan *precision* yang lebih tinggi, menunjukkan bahwa ringkasan lebih akurat dalam memilih kata kunci. Nilai *recall* yang rendah menunjukkan bahwa sistem belum sepenuhnya berhasil menangkap semua informasi penting dari teks asli. Nilai *F1-Score* yang berada pada kategori sedang, menunjukkan bahwa sistem memiliki kemampuan yang cukup baik dalam mengekstraksi informasi penting dari artikel, meskipun masih terdapat beberapa kelemahan. Skor tersebut menunjukkan bahwa sistem masih memiliki kemungkinan untuk memasukkan informasi yang kurang relevan dalam ringkasan

atau justru mengabaikan informasi penting yang seharusnya dimasukkan pada ringkasan. Sementara itu, nilai akurasi yang cukup tinggi menandakan bahwa model mampu menghasilkan ringkasan yang cukup relevan dengan isi dokumen sumber.

Hasil pengujian secara keseluruhan menjelaskan bahwa sistem peringkasan teks artikel jurnal ilmiah bahasa indonesia menggunakan *Fuzzy C-Means* dan *Vector Space Model* memiliki nilai performa yang cukup baik dalam mengekstrak informasi tepat dari isi artikel. Sistem mampu menghasilkan ringkasan teks ekstraktif yang akurat dan mengandung informasi yang relevan.

5.2 Saran

Pengujian yang telah dilakukan untuk sistem peringkasan teks otomatis artikel jurnal ilmiah masih banyak kekurangan di dalam penelitian ini. Model yang dihasilkan belum cukup dikatakan optimal karena belum mampu menyeimbangkan antara nilai *precision* dan *recall.* Berikut beberapa saran untuk peneliti selanjutnya supaya bisa menghasilkan keseimbangan nilai *precision* dan *recall*:

1. Menggunakan model yang berbasis Transformer, yang memiliki kemampuan lebih canggih dalam memahami konteks dan memilih informasi yang relevan.
2. Menggunakan referensi peringkasan manual yang berkualitas baik, karena kualitas ringkasan referensi juga sangat berpengaruh terhadap model.

DAFTAR PUSTAKA

Adelia, R., Suyanto, S., & Wisesty, U. N. (2019). Indonesian abstractive text summarization using bidirectional gated recurrent unit. *Procedia Computer Science*, *157*, 581–588. https://doi.org/10.1016/j.procs.2019.09.017

Aditya, C. S. K., & Wiyono, B. S. (2023). Pengembangan Fitur Peringkasan Artikel Otomatis pada Media Online Satukanal. *BAKTIMAS: Jurnal Pengabdian Pada Masyarakat*, *5*(1), 60–67.

Afsharizadeh, M., Ebrahimpour-Komleh, H., & Bagheri, A. (2018). Query-oriented text summarization using sentence extraction technique. *2018 4th International Conference on Web Research, ICWR 2018*, 128–132. https://doi.org/10.1109/ICWR.2018.8387248

Baeza-yates, R., & Ribeiro-neto, B. (n.d.). The Concepts and Technology behind Search - 15 Enterprise Search. *Modern Information Retrieval*.

Bharathi, R., Shirwaikar, S. C., & Kharat, V. (2016). A distributed, scalable parallelization of fuzzy c-means algorithm. *IEEE Bombay Section Symposium 2016: Frontiers of Technology: Fuelling Prosperity of Planet and People, IBSS 2016*. https://doi.org/10.1109/IBSS.2016.7940196

Dewi, K. E., & Widiastuti, N. I. (2022). The Design of Automatic Summarization of Indonesian Texts Using a Hybrid Approach. *Jurnal Teknologi Informasi Dan Pendidikan*, *15*(1), 37–43. https://doi.org/10.24036/jtip.v15i1.451

Febriani, A. R. (2024). *Sebaik-baik Manusia Adalah yang Bermanfaat bagi Orang Lain, Ini Haditsnya Baca artikel detikhikmah, "Sebaik-baik Manusia Adalah yang Bermanfaat bagi Orang Lain, Ini Haditsnya."* https://www.detik.com/hikmah/doa-dan-hadits/d-7123193/sebaik-baik-manusia-adalah-yang-bermanfaat-bagi-orang-lain-ini-haditsnya#

Gunawan, G., Fitria, F., Setiawan, E. I., & Fujisawa, K. (2023). Maximum Marginal Relevance and Vector Space Model for Summarizing Students' Final Project Abstracts. *Knowledge Engineering and Data Science*, *6*(1), 57. https://doi.org/10.17977/um018v6i12023p57-68

Halimah, Surya Agustian, & Siti Ramadhani. (2022). Peringkasan teks otomatis (automated text summarization) pada artikel berbahasa indonesia menggunakan algoritma lexrank. *Jurnal CoSciTech (Computer Science and Information Technology)*, *3*(3), 371–381. https://doi.org/10.37859/coscitech.v3i3.4300

Irfan, M., Jumadi, Zulfikar, W. B., & Erik. (2017). Implementation of Fuzzy C-Means algorithm and TF-IDF on English journal summary. *Proceedings of the 2nd International Conference on Informatics and Computing, ICIC 2017*, *2018-Janua*, 1–5. https://doi.org/10.1109/IAC.2017.8280646

Jatmiko, W. (2015). *Fakultas Ilmu Komputer Universitas Indonesia PENULISAN ARTIKEL ILMIAH Panduan*.

Johra, M. B. (2021). Soft Clustering Dengan Algoritma Fuzzy K-Means (Studi Kasus : Pengelompokan Desa Di Kota Tidore Kepulauan). *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, *15*(2), 385–392. https://doi.org/10.30598/barekengvol15iss2pp385-392

Lamba, M., & Madhusudhan, M. (2022). *Text Pre-Processing BT - Text Mining for Information Professionals: An Uncharted Territory* (M. Lamba & M. Madhusudhan (eds.); pp. 79–103). Springer International Publishing. https://doi.org/10.1007/978-3-030-85085-2_3

Lin, C.-Y. (2004). ROUGE: A Package for Automatic Evaluation of Summaries Chin-Yew. *Japanese Circulation Journal*, *34*(12), 1213–1220. https://doi.org/10.1253/jcj.34.1213

Nurjanah, Andi Farmadi, F. I. (2014). *Implementasi Metode Fuzzy C-Means Pada Sistem Clustering Data Varietas Padi*. *01*(01), 23–32.

Phu, V. N., Dat, N. D., Ngoc Tran, V. T., Ngoc Chau, V. T., & Nguyen, T. A. (2017). Fuzzy C-means for english sentiment classification in a distributed system. *Applied Intelligence*, *46*(3), 717–738. https://doi.org/10.1007/s10489-016-0858-z

Salma. (2022). *Artikel Ilmiah: Pengertian, Fungsi, Ciri-ciri dan Sistematika*. Deepublish. https://penerbitdeepublish.com/pengertian-artikel-ilmiah/#1_Komara_2017

Samosir, F. V. P., Toba, H., & Ayub, M. (2022). BESKlus : BERT Extractive Summarization with K-Means Clustering in Scientific Paper. *Jurnal Teknik Informatika Dan Sistem Informasi*, *8*(1). https://doi.org/10.28932/jutisi.v8i1.4474

Setiawan, A. Y., Darmawiguna, I. G. M., & Pradnyana, G. A. (2022). Sentiment Summarization Evaluasi Pembelajaran Menggunakan Algoritma LSTM (long short term memory). *Kumpulan Artikel Mahasiswa Pendidikan Teknik Informatika (KARMAPATI)*, *11*(2), 183–191.

Singh, R., & Singh, S. (2021). Text Similarity Measures in News Articles by Vector Space Model Using NLP. *Journal of The Institution of Engineers (India): Series B*, *102*(2), 329–338. https://doi.org/10.1007/s40031-020-00501-5

Singh, V. K., Tiwari, N., & Garg, S. (2011). Document clustering using K-means, heuristic K-means and fuzzy C-means. *Proceedings - 2011 International Conference on Computational Intelligence and Communication Systems, CICN 2011*, 297–301. https://doi.org/10.1109/CICN.2011.62

Suwija Putra, I. M., Adiwinata, Y., Singgih Putri, D. P., & Sutramiani, N. P. (2021). Extractive Text Summarization of Student Essay Assignment Using Sentence Weight Features and Fuzzy C-Means. *International Journal of Artificial*

Intelligence Research, *5*(1), 13–24. https://doi.org/10.29099/ijair.v5i1.187

Utomo, M. S., Wibowo, J. S., & Wahyudi, E. N. (2022). Text Summarization Pada Artikel Berita Menggunakan Vector Space Model Dan Cosine Similarity. *Jurnal Dinamika Informatika*, *14*(1), 11–24. https://doi.org/10.35315/informatika.v14i1.9163

Verma, P., Pal, S., & Om, H. (2019). A comparative analysis on Hindi and English extractive text summarization. *ACM Transactions on Asian and Low-Resource Language Information Processing*, *18*(3). https://doi.org/10.1145/3308754

Wahyono. (2023). *Jumlah Publikasi Jurnal Ilmiah Terindeks Scopus asal Indonesia Tahun ke Tahun, Naik Atau Turun?* SindoNews.Com. https://edukasi.sindonews.com/read/1195439/211/jumlah-publikasi-jurnal-ilmiah-terindeks-scopus-asal-indonesia-tahun-ke-tahun-naik-atau-turun-1694063382

Zamzam, M. A. (2020). Sistem Automatic Text Summarization Menggunakan Algoritma Textrank. *Matics*, *12*(2), 111–116. https://doi.org/10.18860/mat.v12i2.8372

LAMPIRAN

LAMPIRAN

Lampiran 1 Data 100 dokumen artikel jurnal ilmiah

Id	Judul
1	Penerapan Algoritma Support Vector Machine (Svm) Dengan Tf-Idf N-Gram Untuk Text Classification
2	Sistem Diagnosa Penyakit Liver Menggunakan Metode Artificial Neural Network: Studi Berdasarkan Dataset Indian Liver Patient Dataset
3	Klasifikasi Tingkat Kerusakan Sektor Pasca Bencana Alam Menggunakan Metode Multimoora Berbasis Web
4	Pemanfaatan Metode Vector Space Model Dan Metode Cosine Similarity Pada Fitur Deteksi Hama Dan Penyakit Tanaman Padi
5	Rekomendasi Merk Mobil Untuk Calon Pembeli Menggunakan Algoritma Decision Tree
6	Klasifikasi Penyakit Diabetes Menggunakan Algoritma Decision Tree
7	Klasifikasi Ulasan Fasilitas Publik Menggunakan Metode Naïve Bayes Dengan Seleksi Fitur Chi-Square
8	Peringkasan Teks Multi-Dokumen Berbahasa Indonesia Dengan Sentence Scoring Dan Svm
9	Analisis Perbandingan Kecepatan Algoritma Selection Sort Dan Bubble Sort
10	Uji Performa Prediksi Gempa Bumi Di Jawa Timur Dengan Artificial Neural Network
11	Analyzing The Effectiveness Of Collaborative Filtering And Content-Based Filtering Methods In Anime Recommendation Systems
12	Peringkasan Multi Dokumen Berbahasa Indonesia Menggunakan Metode Recurrent Neural Network
13	Optimasi Konten Pemasaran Dan Platform Online Dengan Teknik Search Engine Optimization
14	Klasifikasi Jenis Kayu Menggunakan Support Vector Machine Berdasarkan Ciri Tekstur Local Binary Pattern
15	Perbandingan Metode Machine Learning Dalam Analisis Sentimen Twitter
16	Penerapan Algoritma A* Sebagai Sistem Pencarian Rute Pada Npc Game Labirin
17	Penerapan Algoritma Term Frequency-Inverse Document Frequency (Tf-Idf) Untuk Text Mining
18	Metode Pose To Pose Untuk Membuat Animasi 3 Dimensi Islami "Keutamaan Berbuka Puasa"
19	Implementasi Algoritma Fuzzy Sugeno Pada Game Pious Boy
20	Pengukuran Kemiripan Makna Kalimat Dalam Bahasa Indonesia Menggunakan Metode Path
21	Analisis Perbandingan Metode Alpha Miner, Inductive Miner Dan Causal-Net Mining Dalam Proses Mining
22	Dekomposisi Model Proses Bisnis Tebang Muat Angkut (Tma) Menggunakan Refined Process Structure Tree (Rpst) Dan Metrik Kompleksitas
23	Peringkasan Teks Otomatis Secara Ekstraktif Pada Artikel Berita Kesehatan Berbahasa Indonesia Dengan Menggunakan Metode Latent Semantic Analysis
24	Peran Komitmen Organisasi Dalam Memediasi Job Satisfaction Dan Job Insecurity Terhadap Turnover Intention
25	Perbandingan Metode Ensemble Learning Pada Klasifikasi Penyakit Diabetes
26	Analisis Kesalahan Bahasa Arab Dalam Percakapan Sehari-Hari Anggota Lembaga Raudlah Al-Lughah Al-Arabiyah Pondok Pesantren Annuqyah Sumenep Madura
27	Pengaruh Pemasaran Digital, Product Innovation, Kolaborasi Brand Dan Ulasan Pelanggan Online Terhadap Purchasing Decision Sepatu Aerostreet

28	Vlog Sebagai Hasil Produk Belajar Siswa Dalam Keterampilan Berbicara Bahasa Arab Di Mtsn Kota Batu Malang
29	Pengaruh Islamic Leadership, Budaya Organisasi Terhadap Kinerja Melalui Motivasi Pada Pengurus Pondok Pesantren Sabilul Muttaqin Kota Mojokerto
30	Homonim Kosakata Bugis Dalam Falsafah Hidup Masyarakat Bugis
31	Seleksi Fitur Gain Ratio Pada Analisis Sentimen Kebijakan Pemerintah Mengenai Pembelajaran Jarak Jauh Dengan K-Nearest Neighbor
32	Klasifikasi Data Forum Dengan Menggunakan Metode Naïve Bayes Classifier
33	Implementasi Data Mining Dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako
34	Peringkasan Otomatis Makalah Menggunakan Maximum Marginal Relevance
35	Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5
36	Analisis Komparasi Algoritma Klasifikasi Data Miningdalam Klasifikasi Website Phishing
37	Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi
38	Klasifikasi Motif Batik Menggunakan Convolutional Neural Network.
39	Analisis Sentimen Respon Masyarakat Terhadap Kabar Harian Covid-19 Pada Twitter Kementerian Kesehatan Dengan Metode Klasifikasi Naive Bayes
40	Klasifikasi Penyakit Gigi Dan Mulut Menggunakan Metode Support Vector Machine
41	Peringkasan Dan Support Vector Machine Pada Klasifikasi Dokumen
42	Algoritma Klasifikasi Data Mining Untuk Memprediksi Siswa Dalam Memperoleh Bantuan Dana Pendidikan
43	Perancangan Sistem Klasifikasi Penyakit Jantung Mengunakan Naïve Bayes
44	Klasifikasi Kelompok Umur Manusia Berdasarkan Analisis Dimensi Fraktal Box Counting Dari Citra Wajah Dengan Deteksi Tepi Canny
45	Klasifikasi Berita Online Dengan Menggunakan Pembobotan Tf-Idfdan Cosine Similarity
46	Klasifikasi Tindak Pidana Hudud Dan Sanksinya Dalam Perspektif Hukum Islam
47	Perbandingan Kinerja Word Embeddingword2Vec, Glove, Dan Fasttext Pada Klasifikasi Teks
48	Implementasi Metode Weight Product Dalam Penentuan Klasifikasi Kelas Tunagrahita
49	Decision Tree Dan Adaboost Pada Klasifikasi Penerima Program Bantuan Sosial
50	Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus
51	Implementasi Model Pengembangan Sistem Gdlc Dan Algoritma Linear Congruential Generator Pada Game Puzzle
52	Algoritma Support Vector Machine (Svm) Untuk Klasifikasi Ekonomi Penduduk Penerima Bantuan Pemerintah Di Kecamatan Simpang Raya Sulawesi Tengah
53	Analisa Keamanan E-Commerce Menggunakan Metode Aes Algoritma
54	Algoritma Maze Generator Recursive Backtracking Untuk Membuat Prosedural Labirin Pada Game Petualangan Labirin 3D
55	Analisis Emosi Wisatawan Menggunakan Metode Lexicon Text Analysis
56	Analisis Perbandingan Algoritma Decision Tree, Knn, Dan Naive Bayes Untuk Prediksi Kesuksesan Start-Up
57	Analisis Sentimen Pengguna Aplikasi Marketplace Tokopedia Pada Situs Google Play Menggunakan Metode Support Vector Machine (Svm), Naïve Bayes, Dan Logistic Regression
58	Analisis Sentimen Perpindahan Ibu Kota Negara Pada Aplikasi Tiktok Menggunakan Metode Lstm
59	Analisis Sentimen Terhadap Permendikbud No.30 Pada Media Sosial Twitter Menggunakan Metode Naive Bayes Dan Lstm
60	Analisis Sentimen Ulasan Pengguna Aplikasi Google Play Menggunakan Naïve Bayes
61	Analisis Sentimen Customer Feedback Tokopedia Menggunakan Algoritma Naïve Bayes

62	Penerapan Fuzzy Mamdani Untuk Sistem Pendukung Keputusan Pemilihan Laptop
63	Sistem Pendukung Keputusan Penerimaan Pegawai Dengan Metode Weighted Product Berbasis Web
64	Klasifikasi Berita Hoax Dengan Menggunakan Metode Naive Bayes
65	Peringkasan Teks Otomatis (Automated Text Summarization) Pada Artikel Berbahasa Indonesia Menggunakan Algoritma Lexrank
66	Implementasi Metode Saw Dalam Sistem Pendukung Keputusan Pemilihan Model Social Customer Relationship Management
67	Penerapan Metode Support Vector Machine (Svm) Dalam Klasifikasi Kualitas Pengelasan Smaw (Shield Metal Arc Welding)
68	Pengembangan Aplikasi Manajemen Sewa Motor Berbasis Progressive Web Apps Di Arfand Motorent
69	Pemanfaatan Metode Vector Space Model Dan Metode Cosine Similarity Pada Fitur Deteksi Hama Dan Penyakit Tanaman Padi
70	Metode Algoritma Support Vector Machine (Svm) Linier Dalam Memprediksi Kelulusan Mahasiswa
71	Rancang Bangun Pengolahan Pendapatan Jasa Handling Airport
72	Penerapan Sistem Pakar Untuk Diagnosa Autis Dengan Metode Forward Chaining
73	Rancang Bangun Sistem Informasi Akademik Berbasis Web Pada Madrasah Aliyah Attaqwa Tangerang
74	Aplikasi Penyaluran Bibit Perkebunan Berbasis Web Pada Dinas Perkebunan Kabupaten Pasaman Barat
75	Pengembangan Library Sistem Pemeringkat Otomatis Berbasis Kata Sifat
76	Perbandingan Akurasi Dan Waktu Proses Algoritma K-Nn Dan Svm Dalam Analisis Sentimen Twitter
77	Sistem Informasi Kepegawaian Upt Kesatuan Pengelolaan Hutan Produksi Kapuas Tengah Unit Xi
78	Analisis Perbandingan Metode Topsis Dan Saw Pada Penilaian Karyawan (Studi Kasus : Pt Pura Barutama Unit Paper Mill 5, 6, 9)
79	Analisis Transportasi Pengangkutan Sampah Di Kota Medan Menggunakan Dynamic Programming
80	Analisis Perbandingan Algoritma C4.5 Dan Naïve Bayes Dalam Memprediksi Penyakit Cerebrovascular
81	Analisa Hasil Perbandingan Poly Kernel Dan Normalisasi Poly Kernel Pada Support Vector Machine Sebagai Metode Klasifikasi Citra Tanda Tangan
82	Implementasi Deep Learning Untuk Optimasi Slump Menggunakan Convolutional Neural Network Pada Pt. Handaru Wijaya Mulya
83	Peningkatan Kualitas Layanan Warga Kelurahan Duri Kepa Dengan Aplikasi Lingkoe
84	Perbandingan Fungsi Optimasi Neural Network Dalam Klasifikasi Kelayakan Calon Suami
85	Implementasi Algoritma Naïve Bayes Untuk Klasifikasi Kesegaran Buah Jeruk
86	Penerapan Algoritma Genetika Dalam Penjadwalan Mata Pelajaran
87	Penerapan Business Intelligence Untuk Analisis Kematian Di Indonesia Tahun 2000-2022
88	Analisis Sentimen Publik Pada Media Sosial Twitter Terhadap Tiket.Com Menggunakan Algoritma Klasifikasi
89	Implementasi K-Medoids Dalam Pengelompokan Fasilitas Pelayanan Kesehatan Pada Kasus Tuberculosis
90	Pengelompokan Kasus Tuberculosis Dengan Algoritma KMeans Berdasarkan Kelurahan Di Kota Bogor
91	Sistem Informasi Communication Book Berbasis Web (Studi Kasus : Sd Salman Al Farisi Bandung)
92	Sistem Penunjang Keputusan Untuk Menentukan Status Gizi Balita Menggunakan Metode Fuzzy Tsukamoto

93	Pengaruh Penggunaan Media Sosial Terhadap Prestasi Belajar Siswa Sma N 3 Depok
94	Klasifikasi Berita Indonesia Menggunakan Metode Naive Bayesian Classification Dan Support Vector Machine Dengan Confix Stripping Stemmer
95	Implementasi Pengenal Tulisan Tangan Menggunakan Optical Character Recognition Dengan Metode Cnn Dan Rnn Pada Dokumen Resi Dan Kuitansi
96	Pemilahan Sampah Menggunakan Model Klasifikasi Support Vector Machine Gabungan Dengan Convolutional Neural Network
97	Perbandingan Akurasi Algoritma Adaboost Dan Algoritma Lightgbm Untuk Klasifikasi Penyakit Diabetes
98	Sistem Perhitungan Kendaraan Menggunakan Algoritma Yolov5 Dan Deepsort
99	Klasifikasi Penentuan Pengajuan Kartu Kredit Menggunakan K-Nearest Neighbor
100	Perancangan Sistem Klasifikasi Penyakit Jantung Mengunakan Naïve Bayes