TUGAS AKHIR

SISTEM PENDETEKSI KEMIRIPAN DOKUMEN *TEXT* MENGGUNAKAN METODE *COSINE SIMILARITY* DAN *JACCARD SIMILARITY*

NARDIANAS SILALAHI

NPM: 193510588

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS ISLAM RIAU

PEKANBARU

2024

HALAMAN PENGESAHAN TUGAS AKHIR

Nama : Nardianas Silalahi

NPM : 193510588

Kelompok Keahlian : Artificial Intelligent

Program Studi : Teknik Informatika

Jenjang Pendidikan : Strata Satu (S1)

Judul TA : Sistem Pendeteksi Kemiripan Dokumen Text Menggunakan Metode Cosine Similarity Dan Jaccard Similarity

Format sistematika dan pembahasan materi pada masing-masing bab dan sub bab dalam tugas akhir ini telah dipelajari dan dinilai relatif telah memenuhi ketentuan-ketentuan dan kriteria- kriteria dalam metode penelitian ilmiah. Oleh karena itu tugas akhir ini dinilai layak dapat disetujui untuk disidangkan dalam ujian **Seminar Tugas Akhir.**

Pekanbaru, 26 November 2023

Di sahkan oleh :

Penguji I	Penguji II
Ana Yulianti, ST.,M.Kom	**Ause Labellapansa, S.T., M.Kom., M.Cs**

Ketua Program Studi Teknik Informatika	Dosen Pembimbing
Dr. Apri Siswanto, S.Kom.,M.Kom	**Dr. Arbi Haza Nasution, M.IT**

HALAMAN PENGESAHAN

DEWAN PENGUJI TUGAS AKHIR

Nama : Nardianas Silalahi

NPM : 193510588

Kelompok Keahlian : Artificial Intelligence

Program Studi : Teknik Informatika

Jenjang Pendidikan : Strata Satu (S1)

Judul TA :Sistem Pendeteksi Kemiripan Dokumen Text Menggunakan Metode Cosine Similarity Dan Jaccard Similarity

Tugas Akhir ini secara keseluruhan dinilai telah memenuhi ketentuan-ketentuan dan kaidah-kaidah dalam penulisan penelitian ilmiah serta telah diuji dan dapat dipertahankan dihadapan dewan penguji. Oleh karena itu, Tim Penguji Ujian Tugas Akhir Fakultas Teknik Universitas Islam Riau menyatakan bahwa mahasiswa yang bersangkutan dinyatakan Telah Lulus Mengikuti Ujian Tugas Akhir Pada Tanggal **31 Januari 2024** dan disetujui serta diterima untuk memenuhi salah satu syarat guna memperoleh gelar Sarjana Strata Satu Bidang Ilmu Teknik Informatika.

Pekanbaru, 7 Februari 2024

Dewan Penguji

1. Pembimbing : Dr. Arbi Haza Nasution,B.IT, M.IT ( )
2. Penguji 1 : Ana Yulianti, ST.,M.Kom ( )
3. Penguji 2 : Ause Labellapansa, S.T., M.Kom., M.Cs ( )

Disahkan Oleh :

Ketua Program Studi
Teknik Informatika

Dr. Apri Siswanto,S.Kom.,M.Kom
NIDN.1016048502

PERNYATAAN KEASLIAN TUGAS AKHIR

Dengan ini saya menyatakan bahwa tugas akhir ini merupakan karya saya sendiri dan semua sumber yang tercantum didalamnya baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar sesuai ketentuan. Jika terdapat unsur penipuan atau pemalsuan data maka saya bersedia dicabut gelar yang telah saya peroleh.

Pekanbaru, 31 Januari 2024

NARDIANAS SILALAHI
NPM: 193510588

KATA PENGANTAR

Puji syukur kehadirat Tuhan Yang Maha Esa yang telah melimpahkan segala rahmat dan karunia Nya kepada penulis, sehingga penulis dapat menyelesaikan proposal skripsi yang berjudul "**Sistem Pendeteksi Kemiripan Dokumen Text Menggunakan Metode *Cosine Similarity* Dan *Jaccard Similarity***" sebagai salah satu syarat wajib untuk mendapatkan gelar sarjana pada Fakultas Teknik Program Studi Informatika Universitas Islam Riau.

Dalam penyusunan proposal skripsi ini, penulis menyadari bahwa penulisan proposal skripsi ini masih jauh dari kata sempurna dan banyak mengalami kendala. Namun, penulis mendapat banyak sekali bantuan, dorongan dan bimbingan dari berbagai pihak. Untuk itu, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada :

1. Orang tua tercinta Bapak Dorlan Silalahi dan Ibu Ermida Hutabarat serta keluarga yang senantiasa memberikan motivasi, dukungan, pengorbanan serta doa yang tidak pernah putus.
2. Bapak Dr. Eng. Muslim, ST., MT selaku Dekan Fakultas Teknik Universitas Islam Riau.
3. Bapak Dr. Apri Siswanto, S.Kom., M.Kom selaku Ketua Program Studi Teknik Informatika.
4. Ibu Ana Yulianti, ST., M.Kom selaku Sekretaris Program Studi Teknik Informatika.
5. Ibu Nesi Syafitri, S.Kom., M.Cs selaku dosen PA yang telah memberikan masukan dan bimbingan selama melaksanakan perkuliahan

6. Bapak Dr Arbi Haza Nasution B.IT.(Hons), M.IT selaku dosen pembimbing yang sangat banyak membantu, membimbing dan memberikan arahan sehingga penulis dapat menyelesaikan laporan skripsi ini dengan baik dan benar.
7. Ibu Ana Yulianti, ST., M.Kom dan Ibu Ause Labellapansa, ST., M.Cs., M.Kom selaku dosen penguji yang telah memberikan masukan dan arahan dalam membuat laporan skripsi ini.
8. Seluruh dosen Program Studi Teknik Informatika Universitas Islam Riau yang telah mendidik dan memberi arahan selama dibangku kuliah.
9. Teman-teman seperjuangan AhSudalah sekaligus keluarga kedua penulis, Syarifah Kusuma Maharani, Ummul Muthmainnah Ulya, Juwita Novi Maradika, Rezky Byon, Rizky Hendriawan, Kholilurohman, Ferdi Alfarabi, dan Eet Mendra, yang telah memberikan dukungan, motivasi serta berbagai sumbangan pemikiran kepada penulis.
10. Seluruh pihak yang tidak dapat disebutkan satu persatu yang telah banyak membantu dan memberikan pemikiran demi kelancaran dan keberhasilan penyusunan skripsi ini.

Penulis menyadari masih banyak kekurangan dalam penyusunan laporan skripsi ini, untuk itu dengan segala kerendahan hati penulis mengharapkan saran dan kritik yang sifatnya membangun dari pembaca untuk penyempurnaan laporan iii skripsi ini. Akhir kata, semoga laporan skripsi ini dapat menambah pengetahuan dan bermanfaat bagi para pembaca.

Pekanbaru ,16 Maret 2024

Penulis,

Nardianas Silalahi

SISTEM PENDETEKSI KEMIRIPAN DOKUMEN *TEXT* MENGGUNAKAN METODE *COSINE SIMILARITY* DAN *JACCARD SIMILARITY*

NARDIANAS SILALAHI

Program Studi Teknik Informatika, Fakultas Teknik, Universitas Islam Riau Email: nardianas@student.uir.ac.id

ABSTRAK

Peningkatan pertukaran informasi melalui dokumen teks dalam era informasi saat ini menimbulkan tantangan dalam pengelolaan dan analisis besar kesamaan dokumen. Plagiarisme, terutama dalam konteks akademis, menjadi risiko signifikan. Penelitian ini fokus pada analisis performa algoritma *Cosine* similarity dan *Jaccard* similarity dalam mendeteksi kemiripan abstrak proposal penelitian. Meskipun keduanya memiliki kelebihan masing-masing, penelitian ini mencari pemahaman lebih baik tentang kapan dan bagaimana metode-metode ini sebaiknya digunakan. Hasil pengujian menunjukkan perbedaan hasil yang relatif kecil antara kedua metode, dengan selisih rata-rata sekitar 3%. Waktu pemrosesan data juga menunjukkan bahwa rata-rata selisih antara *Jaccard* similarity dan *Cosine* similarity tidak terlau jauh dan cendrung hampir sama dengan rata-rata selisih 0.28 detik dengan *Cosine* similarity lebih cepat, selisihnya kecil dan dapat diterima. meskipun terdapat perbedaan hasil, keduanya layak diimplementasikan dalam Sistem Purse Universitas Islam Riau untuk meningkatkan efisiensi, akurasi, dan integritas dalam proses penelitian.

Kata kunci: teks *mining*, *Cosine similarity*, *Jaccard similariry*.

TEXT DOCUMENT SIMILARITY DETECTION SYSTEM USING *COSINE* SIMILARITY AND *JACCARD* SIMILARITY METHODS

NARDIANAS SILALAHI

Informatics Engineering Study Program, Faculty of Engineering, Riau Islamic University

Email: nardianas@student.uir.ac.id

ABSTRACT

The increased exchange of information through text documents in today's information age poses challenges in the management and analysis of large document similarities. Plagiarism, especially in an academic context, becomes a significant risk. This research focuses on analyzing the performance of *Cosine* similarity and *Jaccard* similarity algorithms in detecting the similarity of research proposal abstracts. While both have their merits, this research seeks a better understanding of when and how these methods should be used. The test results showed a relatively small difference in results between the two methods, with an average difference of about 3%. The data processing time also shows that the average difference between *Jaccard* similarity and *Cosine* similarity is not too far and tends to be almost the same with an average difference of 0.28 seconds with *Cosine* similarity being faster, the difference is small and acceptable. despite the difference in results, both are worth implementing in the Riau Islamic University Purse System to improve efficiency, accuracy, and integrity in the research process.

Keywords: text mining, Cosine similarity, Jaccard similarity,

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

BAB I

PENDAHULUAN

1.1 Latar Belakang

Peningkatan signifikan dalam penggunaan dan pertukaran informasi dalam bentuk dokumen teks telah menjadi ciri khas dari era informasi saat ini. Dengan semakin meluasnya akses ke sumber daya digital dan platform berbagi, muncul tantangan baru terkait dengan pengelolaan, analisis, dan pemrosesan besar volume dokumen teks. Dalam konteks ini, penting untuk mengembangkan alat yang dapat mengidentifikasi dan menganalisis kemiripan antara dokumen teks untuk berbagai tujuan seperti penghapusan redundansi, penemuan informasi, dan penerapan hukum hak cipta.

Menurut Subroto & Selamat (2019) kemajuan teknologi mendukung pesatnya distribusi dokumen karya ilmiah secara publik sebagai referensi sebuah riset. Namun, luasnya distribusi dan kemudahan akses ini dapat menjadi celah terjadinya tindakan plagiarisme. Pada akademisi sanksi administratif yang dapat diberikan atas tindakan plagiarisme adalah teguran hingga yang paling berat pembatalan ijazah dan pemberhentian secara tidak hormat dari jabatan yang diduduki berdasarkan PERMENDIKNAS No. 17 tahun 2010 tentang Pencegahan dan Penanggulangan Plagiat di Perguruan Tinggi Yulianti (2020). Seiring pertumbuhan data dokumen yang sangat pesat, akan menjadi mustahil dilakukan inspeksi orisinalitas secara manual. Dalam deteksi plagiarisme teks secara literal antar dokumen telah dilakukan dengan berbagai metode, diantaranya yang paling pupuler dilakukan adalah *Jaccard* Coefficient dan *Cosine* Similarity .

Penelitian mengenai pendeteksian kemiripan dokumen teks telah menjadi perhatian utama dalam bidang Teknik Informatika. Tujuan utama dari penelitian ini adalah untuk mengembangkan metode yang efisien dan akurat dalam mengidentifikasi sejauh mana dua atau lebih dokumen teks serupa dalam konten atau topik tertentu. Dua metode yang umum digunakan untuk tujuan ini adalah *Cosine* Similarity dan *Jaccard* Similarity.

Cosine similarity merupakan salah satu metode untuk menentukan kesamaan antara dua objek, *Cosine* Kemiripan menggunakan dua vektor yang menyajikan dua dokumen teks dimana kosinus nilai sudut kedua vektor adalah nilai kemiripan keduanya dokumen teks. Batas nilai yang dihasilkan berkisar antara 0 sampai dengan 1. Susunan kata dalam dokumen teks adalah penentuan nilai yang diperoleh dari *Cosine* Metode kesamaan. Metode *Cosine* Kemiripan tidak bisa menentukan arti umum setiap kata]. Setiap kata yang mempunyai komponen huruf berbeda dianggap kata yang berbeda (Badruzzaman, 2020).

Jaccard Similarity, di sisi lain, adalah metode yang lebih sederhana yang berfokus pada perbandingan himpunan token atau kata-kata yang ada dalam dua dokumen. Ini merupakan metode yang berguna untuk mengidentifikasi kesamaan relatif antara dokumen teks, terutama jika ukuran dokumen lebih kecil (Abhilash, 2023).

Namun, meskipun kedua metode ini memiliki kelebihan masing-masing, tidak selalu jelas metode mana yang lebih cocok dalam berbagai situasi. Dalam konteks Teknik Informatika, pendeteksian kemiripan dokumen teks memiliki potensi aplikasi yang luas. Ini mencakup bidang seperti penggalian data, analisis teks, pemrosesan bahasa alami, dan pemahaman konten. Selain itu,

pengembangan metode yang lebih baik dalam pendeteksian kemiripan dokumen teks akan memberikan kontribusi positif terhadap peningkatan kualitas sistem pengelolaan informasi.

Proses pengecekan plagiasi pada abstrak proposal penelitian menentukan integritas penelitian dan kualitas hasil akhir. Meskipun abstrak sering kali mengandung kutipan penting dan informasi esensial, kurangnya pengecekan awal plagiasi pada tahap ini dapat merugikan, terutama ketika peneliti tidak dilibatkan secara efektif. Oleh karena itu, perlunya menganalisis performa algoritma *Cosine* similarity dan *Jaccard* similarity menjadi krusial dalam memastikan keakuratan dan ketepatan deteksi kemiripan teks.

Pengecekan plagiasi pada proposal penelitian muncul karena praktik ini sering diabaikan atau bahkan tidak dilakukan secara efektif. Kondisi ini membuka celah bagi peneliti untuk mengajukan proposal yang mengandung materi yang telah disalin tanpa atribusi yang tepat, mengancam integritas akademis. Selain itu, pengecekan awal plagiasi pada proposal penelitian belum diimplementasikan di Sistem Purse Universitas Islam Riau, meningkatkan urgensi penelitian ini untuk mengisi kekosongan tersebut.

Penelitian ini akan fokus menganalisis performa algoritma *Cosine* similarity dan *Jaccard* similarity dalam mendeteksi kemiripan abstrak proposal penelitian. Dengan mempertimbangkan kebutuhan mendesak untuk penerapan pengecekan awal plagiasi, terutama di lingkungan Sistem Purse Universitas Islam Riau, penelitian ini diharapkan dapat memberikan landasan yang kokoh untuk peningkatan efisiensi, akurasi, dan integritas dalam proses penelitian.

1.2 Identifikasi Masalah

Berdasarkan latar belakang diatas, dapat diidentifikasi beberapa masalah seperti berikut :

1. Menganalisis performa dari algoritma *Cosine similarity* dan *Jaccard similarity* dalam mendeteksi kesamaan teks antar dokumen.
2. Pentingnya pengecekan plagiasi pada proposal penelitian: pengecekan plagiasi pada proposal penelitian sering kali tidak dilakukan secara efektif atau bahkan tidak dilakukan sama sekali. Hal ini membuka peluang bagi peneliti untuk mengajukan proposal yang mengandung bagian-bagian yang telah disalin dari sumber lain tanpa atribusi yang tepat.
3. Belum adanya Pengecekan Awal Plagiasi pada Proposal Penelitian di Sistem Purse Universitas Islam Riau.

1.3 Rumusan Masalah

Berdasarkan latar belakang masalah di atas, maka dapat dirumuskan masalah berikut :

1. Bagaimana cara mengembangkan sistem pendeteksi kemiripan dokumen teks yang efisien dan akurat?
2. Bagaimana pengaruh penggunaan metode *Cosine* similarity dan *Jaccard* similarity terhadap hasil pendeteksian kemiripan dokumen teks?

1.4 Batasan Masalah

Dalam penelitian ini, obyek penelitian dibatasi dengan ruang lingkup sebagai berikut:

1. Berkas dokumen teks yang dibandingkan adalah berupa dokumen digital dengan ekstensi berkas txt, pdf, dan csv.
2. Berkas dokumen teks yang dibandingkan adalah dokumen teks yang seluruhnya atau sebagian besar.
3. Berkas dokumen teks yang dibandingkan adalah dokumen teks yang memiliki penulisan ejaan yang dibenarkan Pedoman Umum Ejaan Bahasa Indonesia (PUEBI).

1.5 Tujuan Penelitian

Adapun tujuan penelitian ini adalah

1. Mengembangkan Sistem Pendeteksi Kemiripan Dokumen: Tujuan pertama penelitian ini adalah mengembangkan sebuah sistem pendeteksi kemiripan dokumen teks yang efisien dan akurat. Sistem ini diharapkan mampu mengolah sejumlah besar dokumen dengan cepat dan memberikan hasil yang andal dalam mengidentifikasi tingkat kemiripan antara mereka.
2. Menganalisis Performa Metode *Cosine* Similarity dan *Jaccard* Similarity: Tujuan kedua penelitian ini adalah untuk menganalisis kinerja metode *Cosine* similarity dan *Jaccard* similarity dalam pendeteksian kemiripan dokumen. Penelitian ini akan mengevaluasi keefektifan kedua metode tersebut, mengidentifikasi kelebihan dan kekurangan masing-masing, serta memberikan pemahaman yang lebih baik tentang kapan dan bagaimana metode-metode ini sebaiknya digunakan.

1.6 Manfaat Penelitian

Penelitian mengenai sistem pendeteksi kemiripan dokumen menggunakan metode *Cosine* similarity dan *Jaccard* similarity memiliki beberapa manfaat yang signifikan:

1. Mencegah Plagiat dan Pelanggaran Hak Kekayaan Intelektual: Sistem pendeteksi yang efisien dan akurat dapat membantu mencegah kasus plagiat dan pelanggaran hak kekayaan intelektual. Dengan mengidentifikasi dokumen yang memiliki tingkat kemiripan yang tinggi, baik secara sengaja maupun tidak sengaja, sistem ini dapat memberikan perlindungan terhadap karya intelektual yang sah.
2. Mendukung Proses Penelitian dan Akademik yang Etis: Di dunia akademik, penting untuk memastikan keaslian dan orisinalitas karya ilmiah. Sistem ini dapat membantu peneliti, mahasiswa, dan akademisi untuk memastikan bahwa tulisan mereka tidak menyalin atau mengambil ide secara tidak sah dari sumber lain. Ini mendorong budaya penelitian yang etis.
3. Efisiensi dalam Pengolahan Dokumen: Dengan adanya sistem pendeteksi yang efisien, pengolahan dan analisis sejumlah besar dokumen teks menjadi lebih cepat dan lebih mudah. Hal ini dapat membantu organisasi atau lembaga yang memiliki database besar untuk mengelola konten mereka dengan lebih baik.
4. Peningkatan Akurasi dalam Pengukuran Kemiripan: Penelitian ini dapat menghasilkan peningkatan dalam akurasi pengukuran kemiripan antara dokumen-dokumen. Dengan menganalisis metode-metode yang ada,

peneliti dapat mengidentifikasi cara-cara untuk meningkatkan hasil yang lebih andal dan akurat.

5. Basis untuk Pengembangan Lebih Lanjut: Hasil dari penelitian ini dapat menjadi dasar untuk pengembangan lebih lanjut dalam bidang deteksi kemiripan dan pengolahan teks. Metode-metode yang dikembangkan atau dievaluasi dalam penelitian ini dapat dijadikan pijakan untuk penelitian lebih mendalam atau pengembangan teknologi baru.
6. Kontribusi terhadap Pengetahuan Ilmiah: Penelitian ini dapat memberikan kontribusi berharga terhadap pengetahuan ilmiah dalam bidang pengolahan bahasa alami, deteksi kemiripan, dan teknologi informasi. Temuan dan hasil penelitian ini dapat diterbitkan dalam jurnal-jurnal ilmiah, sehingga dapat diakses oleh komunitas akademik dan profesional.

Dengan adanya manfaat-manfaat ini, penelitian mengenai sistem pendeteksi kemiripan dokumen menggunakan metode *Cosine* similarity dan *Jaccard* similarity memiliki potensi untuk memberikan dampak positif dalam berbagai aspek kehidupan, termasuk akademik, industri, dan riset.

BAB II

LANDASAN TEORI

2.1 Tinjauan Pustaka

Penelitian terdahulu ini menjadi salah satu acuan penulis dalam melakukan penelitian sehingga penulis dapat memperkaya teori yang digunakan dalam mengkaji penelitian yang dilakukan. Dari penelitian terdahulu, penulis tidak menemukan penelitian dengan judul yang sama seperti judul penelitian penulis. Namun penulis mengangkat beberapa penelitian sebagai referensi dalam memperkaya bahan kajian pada penelitian penulis. Berikut merupakan penelitian terdahulu berupa beberapa jurnal terkait dengan penelitian yang dilakukan penulis.

Rito Putriwana Pratama, Dkk (2019) membuat penelitian dengan judul "Deteksi Plagiarisme pada Artikel Jurnal Menggunakan Metode *Cosine* Similarity" penelitian ini bertujuan untuk mengembangkan sebuah sistem pendeteksi plagiarisme pada dokumen jurnal online dengan menggunakan metode *Cosine* similarity dan web crawler. Selain itu, penelitian ini juga bertujuan untuk mengumpulkan data dari jurnal online yang akan dijadikan dokumen repository untuk proses deteksi plagiarisme. Hasil dari penelitian ini kemudian akan dianalisis untuk mengevaluasi keakuratan sistem dan metode yang digunakan.

Heri Sutikno, (2021) membuat penelitian dengan judul "Implementasi Algoritma *Cosine* Similarity Untuk Mendeteksi Kemiripan Topik Judul" penelitian ini bertujuan untuk membentuk suatu system yang membantu dalam membandingkan beberapa tugas akhir. Dalam sistem ini judul di input kedalam

sebuah sistem kemudian data akan melewati beberapa tahapan yaitu: Text Minning (Tokenzing, Filtering, Stemming, Stopword Removal). Proses berikutnya yaitu pembobotan TF_IDF dan perhitungan *Cosine* Similarity. Hasil akhir dari proses tersebut adalah tingkat kemiripan antar judul yang di uji.

Sapto Utomo, Dkk (2022) membuat penelitian dengan judul "Deteksi Plagiat Tugas Akhir dengan Metode *Jaccard* Similarity" penelitain ini bertujuan untuk membuat sebuah sistem pendetesi untuk mencegah adanya unsur plagiarisme dalam pembuatan laporan tugas akhir, mahasiswa tidak boleh meniru persis kata atau kalimat yang akan dijadikan sebagai acuan dari pembuatan laporan tugas akhir.

Ginting, Dkk (2022) membuat penelitian dengan judul "Aplikasi Deteksi Kemiripan Kata Menggunakan Algoritma Rabin-Karp" penelitian bertujuan untuk merancang sebuah aplikasi pendeteksi kemiripan kata pada dokumen. Metode yang diterapkan adalah algoritma Rabin-Karp dengan menggunakan pola string, sistem pendeteksi kesamaan kata ini dapat membantumengidentifikasi kesamaan antara dua dokumen yang dibandingkan.

Made, Dkk (2021) memnbuat penelitian denngan judul "Deteksi Kesamaan Teks Jawaban pada Sistem Test Essay Online dengan Pendekatan Neural Network". Tujuan dari penelitian ini adalah mengembangkan dan menerapkan sebuah sistem pendeteksi kesamaan jawaban essay siswa pada platform e-learning, dengan menggunakan metode Artificial Neural Network (ANN), Latent Semantic Index (LSI), dan *Jaccard* Similarity. Sistem ini bertujuan membantu mencegah tindakan plagiarisme antara siswa dalam tugas essay, khususnya dalam hal menilai hasil kegiatan belajar yang kompleks. Metode ANN

dipilih karena memiliki potensi untuk mendeteksi kesamaan teks dengan kinerja yang lebih mendekati metode *Jaccard* dibandingkan dengan metode LSI, seperti yang diuji dan dievaluasi pada penelitian ini. Keseluruhan penelitian ini dilakukan untuk meningkatkan keakuratan dan efektivitas dalam proses evaluasi jawaban essay siswa melalui teknologi e-learning..

2.2 Dasar Teori

2.2.1 Sistem

Sistem adalah suatu jaringan kerja dari prosedur yang saling berhubungan, berkumpul bersama-sama untuk melakukan sasaran tertentu. Sedangkan informasi adalah data yang diolah menjadi bentuk yang lebih berguna dan lebih berarti bagi penerimanya. Sehingga dapat disimpulkan bahwa sistem informasi adalah suatu sistem didalam suatu organisasi yang mempertemukan kebutuhan pengelolaan transaksi harian, mendukung operasi, bersifat manajerial, dan kegiatan strategi dari suatu organisasi tertentu dengan laporan-laporan yang dibutuhkan (Yeni,2020).

2.2.2 *Text Mining*

Text mining merupakan proses ekstraksi pola (informasi dan pengetahuan yang berguna) dari banyaknya data yang tak terstruktur. *Text mining* adalah disiplin ilmu yang mencakup pengambilan informasi (information retrieval), analisa teks, ekstraksi informasi, kategorisasi, pengelompokan (clustering), visualisasi, penambangan data (data mining), dan pembelajaran mesin (machine learning) (Dahniawati et al., 2019).

Text mining merupakan proses mengeksplorasi dan menganalisis data dalam bentuk teks dengan tujuan mengidentifikasi konsep, pola dan kata kunci,

data yang berupa teks dimana sumber data biasanya didapatkan dari dokumen dan tujuannya adalah mencari kata-kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan analisa keterhubungan antar dokumen (Geovaldo,2021).

2.2.3 *Python*

Python adalah bahasa pemrograman interpretatif multiguna. Tidak seperti bahasa lain yang susah untuk dibaca dan dipahami, python lebih menekankan pada keterbacaan kode agar lebih mudah untuk memahami sintaks. Hal ini membuat Python sangat mudah dipelajari baik untuk pemula maupun untuk yang sudah menguasai bahasa pemrograman lain. Bahasa ini muncul pertama kali pada tahun 1991, dirancang oleh seorang bernama Guido van Rossum. Sampai saat ini Python masih dikembangkan oleh Python Software Foundation. Bahasa Python mendukung hampir semua sistem operasi, bahkan untuk sistem operasi Linux, hampir semua distronya sudah menyertakan Python di dalamnya (Anonymous, 2019).

2.2.4 *Naturan Language Processing (NLP)*

Pemrosesan bahasa alami adalah rangkaian teknik komputasi secara teoritis untuk menganalisis dan mempresentasikan teks yang terjadi secara alami pada satu atau lebih tingkat analisis linguistik untuk tujuan mencapai pemrosesan bahasa mirip manusia dalam menyelesaikan beberapa task atau aplikasi (Liddy, 2019). Beberapa istilah yang sering digunakan dalam NLP seperti token, sentence, tokenization, filtering, case folding, stemming, stopword dan corpus.

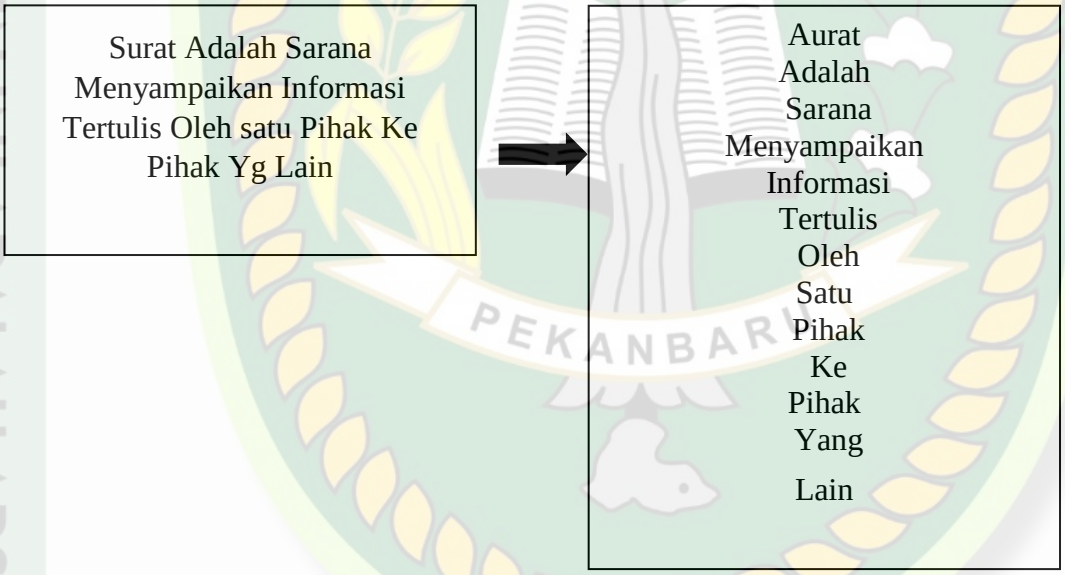
2.2.5 *Natural Language Tool Kit (NLTK)*

NLTK merupakan library khusus pendukung untuk pengolahan text stemming yang menggunakan bahasa pemrograman python. Dengan adanya

library ini seorang peneliti akan dengan mudah melakukan stemming karena hanya perlu memanggil fungsi dan sintax librarynya saja (Ulil Albab, 2023). beberapa pre-processed yang digunakan dari Natural Language Tool Kit digunakan dalam penelitian ini adalah sebagai berikut.

A. *Tokenizing*

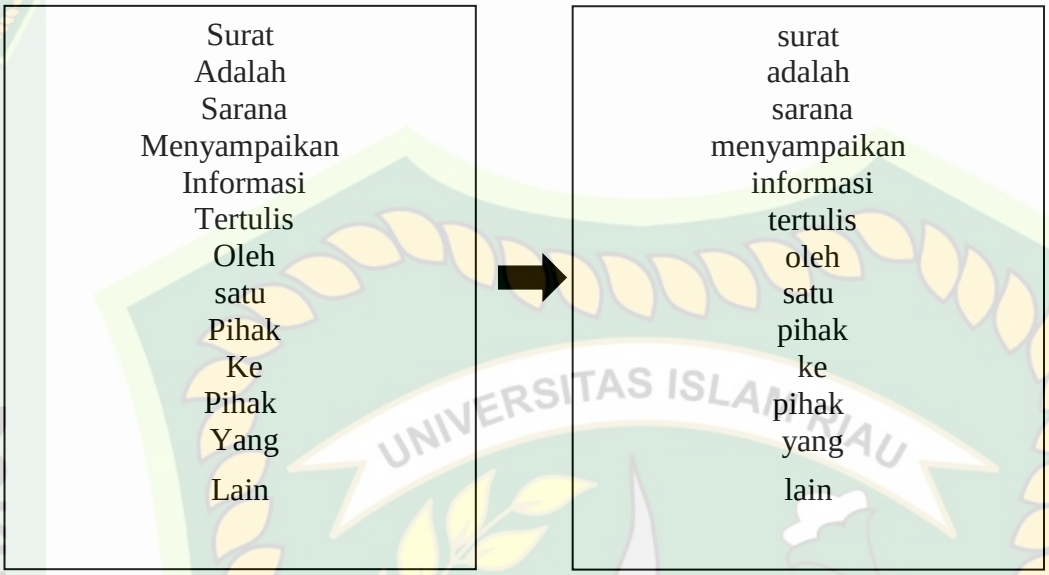
Tokenizing adalah sebuah proses pemotongan kalimat menjadi setiap kata yang menyusunnya. Potongan-potongan kata ini disebut token atau term. Tokenizing memotong tiap kata dalam kalimat menggunakan spasi sebagai delimiter yang akan menghasilkan token (Brata & Hetami, 2019).



Gambar2.1 Contoh Tokenizing

B. **Case Folding**

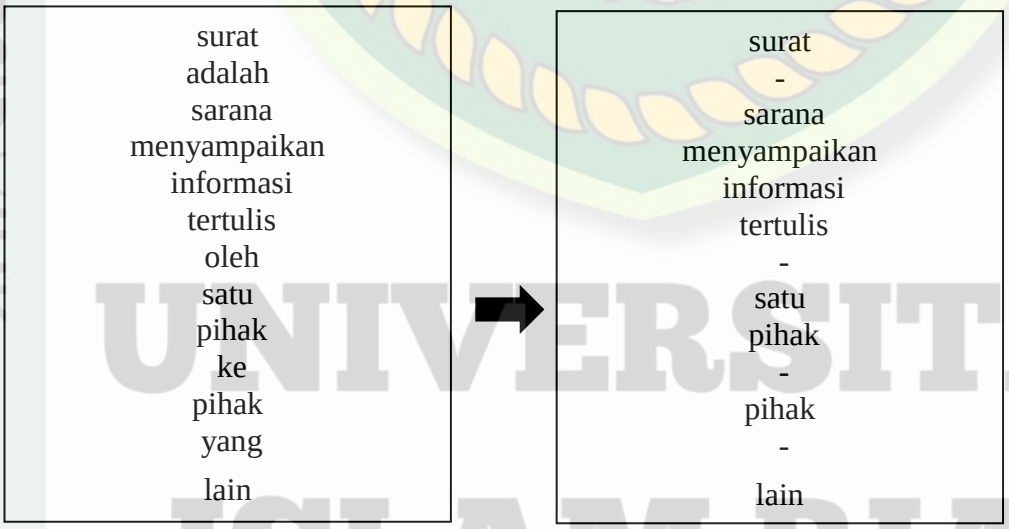
Case folding merupakan langkah konversi font dengan mengubah semua huruf menjadi huruf kecil. Case folding merupakan proses mengubah semua huruf di dalam dokumen menjadi huruf kecil (Brata & Hetami, 2019).



Gambar2.2 Contoh Case Folding

C Stopword Removal

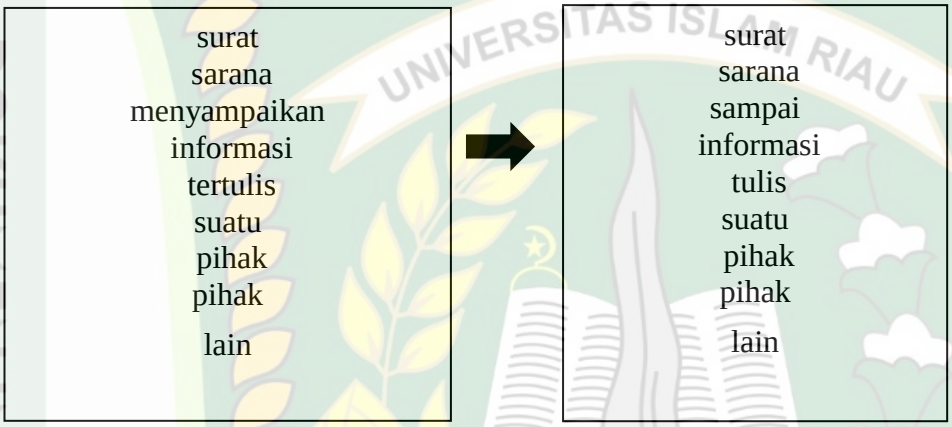
Stopword remooval adalah pengapusan kata-kata umum (Commond words) yang sering muncul, yang tidak memiliki arti atau informasi penting (yang biasanya tidak diacuhkan atau dibuang misalnya dalam proses pembuatan indeks atau daftar kata). Contoh stopword bahasa Indonesia antara lain "yang","di","ke",dan lain-lain (Anwar, 2019).



Gambar 2.3 Contoh stopword removal

D Stemming

Stemming, yaitu proses menemukan kata dasar dari sebuah kata dengan menghapus semua imbuhan (affixes) baik yang terdiri dari awalan (prefixes), sisipan (infixes), akhiran (suffixes) serta kombinasi dari awalan dan akhiran (confixes) pada kata turunan (Febriyanto, 2019).



Gambar2.4 Contoh Stemming

E. Term Frequency Inverse Document Frequency (*TF-IDF*)

Metode TF-IDF merupakan metode untuk menghitung bobot suatu kata (term) terhadap dokumen. Metode ini juga terkenal efisien, mudah dan memiliki hasil yang akurat. Metode ini menggabungkan dua konsep untuk perhitungan bobot, yaitu frekuensi kemunculan sebuah kata didalam sebuah dokumen tertentu dan inverse frekuensi dokumen yang mengandung kata tersebut. Frekuensi kemunculan kata di dalam dokumen yang diberikan menunjukan seberapa penting kata tersebut di dalam dokumen.

2.2.6 *Jaccard Similarity*

Jaccard Algorithm atau dikenal dengan *Jaccard* Coefficient dan atau *Jaccard* Similarity adalah salah satu metode yang dipakai untuk menghitung similarity

antara dua objek (items) . Masing-masing dokumen akan dihitung kata yang sama antara dokumen yang satu dengan dokumen yang lain. Hasil dari perhitungan akan dihasilkan nilai similaritas dokumen. Nilai similaritas dokumen yang tertinggi dapat dianggap bahwa dokumen tersebut paling similar, atau memiliki banyak kesamaan (Sapto Utomo., 2022). Namun *Jaccard* similarity memiliki kelemahan dimana perhitungannya tidak memperhatikan term frequency(berapa kali suatu term terdapat di dalam suatu dokumen). *Jaccard* similarity dapat dirumuskan sebagai berikut:

$$Jaccard\ (X,Y) = \frac{|X \cap Y|}{|X \cup Y|} \ldots (2)$$

Dimana :

X = Surat Adalah Sarana Menyampaikan Informasi Tertulis Oleh satu Pihak Ke Pihak Yang Lain.

Y = Dokumen Adalah Sebuah Tulisan Yang Memuat Informasi.

Lalu setelah melalui tahap preprocessing maka akan menjadi :

X = surat sarana sampai informasi tulis satu pihak pihak lain

Y = dokumen buah tulis muat informasi

Rumus *Jaccard* akan digunakan untuk mencari persamaan dan perbedaan pada dua sampel. Sebagai contoh :

$X \cap Y$ = { surat, sarana, sampai, informasi, tulis, satu, pihak, pihak, lain }

$X \cup Y$ = { dokumen, buah, tulis, muat, informasi }

maka akan menghasilkan nilai:

$$Jaccard\ (X,Y\ ) = \frac{|\ surat,sarana,sampai,informasi,tulis,satu,pihak,pihak,lain\ |}{|\ dokumen,buah,tulis,muat,informasi\ |}$$

$$Jaccard\ (X,Y) = \frac{2}{14} = 0{,}14 = 14\%$$

Berdasarkan nilai persamaan yang diperoleh maka dapat ditetapkan bahwa nilai kemiripan dari dokumen 1 dan dokumen 2 adalah 0,14 atau 14%

2.2.7 *Cosine similarity*

Cosine Similarity adalah metode yang digunakan untuk menentukan seberapa mirip antara dua vector yang diberikan. Metode *Cosine* Similarity digunakan untuk menentukan kesaman antara dua objek yang dimyatakan dalam bentuk vector. *Cosine* Similarity mengukur kesamaan antara dua vector dengan menghitung *Cosine* dari sudut antara kedua vector. (Samosir dan Nurzaman, 2023). *Cosine* similarity dapat dirumuskan sebagai berikut:

$$Similatiry\ (A,B) = = \frac{A.B}{\|A\|.\|B\|} = \frac{\sum_{i=1}^{n} A.B}{\sqrt{\sum_{i=1}^{n}(A)^2}.\sqrt{\sum_{i=1}^{n}(B)^2}}$$

Di mana :

A = Vektor A, yang akan dibandingkan kemiripannya

B = Vektor B, yang akan dibandingkan kemiripannya

A.B = dot product antara vektor A dan vektor B

|A| = panjang vektor A

|B| = panjang vektor B

|A||B| = cross product antara |A| dan |B|

Rumus tersebut digunakan untuk mencari kemiripan antar teks dokumen.

Sebagai contoh:

D1: surat sarana sampai pihak informasi surat tulis satu pihak pihak lain

D2: dokumen buah tulis muat buah informasi

Tabel 2. 1 Pembobotan Kata

Term	D1	D2
Surat	2	0
Sarana	1	0
Sampai	1	0
Informasi	1	1
Tulis	1	1
Satu	1	0
Pihak	3	0
Lain	1	0
Dokumen	0	1
Buah	0	2
Muat	0	1

Vektor D1 = 1, 1, 2, 1 ,1 ,1 ,3 ,1 ,0 ,0 ,0

Vektor D2 = 0, 0, 0, 1, 2, 0, 0, 0, 1, 2, 1

$$= \frac{(1*0)+(1*0)+(1*0)+(1*1)+(1*1)+(1*0)+(3*0)+(1*0)+(0*1)+(0*1)+(0*1)}{\sqrt{1^{2+}\ 1^{2+}\ 2^{2+}\ 1^{2+}\ 1^{2+}\ 1^{2+}\ 3^{2+}\ 0^{2+}\ 0^{2+}\ 0^{2+}\ 0^{2}} * \sqrt{0^{2+}\ 0^{2+}\ 0^{2+}\ 1^{2+}\ 2^{2+}\ 0^{2+}\ 0^{2+}\ 0^{2+}\ 1^{2+}\ 2^{2+}\ 1^{2}}}$$

$$= \frac{0+0+0+1+1+0+0+0+0+0+0}{\sqrt{18} * \sqrt{11}}$$

$$= \frac{2}{14}$$

=0,14

Berdasarkan nilai persamaan yang diperoleh maka dapat ditetapkan bahwa nilai kemiripan dari dokumen 1 dan dokumen 2 adalah 0,14 atau 14%.

BAB III
METODOLOGI PENELITIAN

3.1 Tempat Penelitian

Penelitian dilaksanakan pada semester ganjil tahun akademik 2023/2024 di Jurusan Teknik Informatika, Fakultas Teknik, Universitas Islam Riau yang beralamat di Jl. Jl. Kaharuddin Nasution No.113 Pekanbaru, Riau.

3.2 Alat Dan Data

Adapun alat dan bahan yang digunakan dalam penelitian ini antara lain:

1. Perangkat Keras (Hardware)

Perangkat keras yang digunakan dalam penelitian ini adalah satu unit laptop dengan spesifikasi:

a.Processor : AMD A9 CPU 2.5GHz

b.Installed RAM : 4 GB

c.System type : 64-bit Operating System

d.Operating System : Windows 10

2. Perangkat Lunak (Software)

Perangkat lunak yang digunakan dalam penelitian ini adalah:

a.Jupyter Notebook

b.python

c.Anaconda Navigator

d.Web Browser Google Chrome

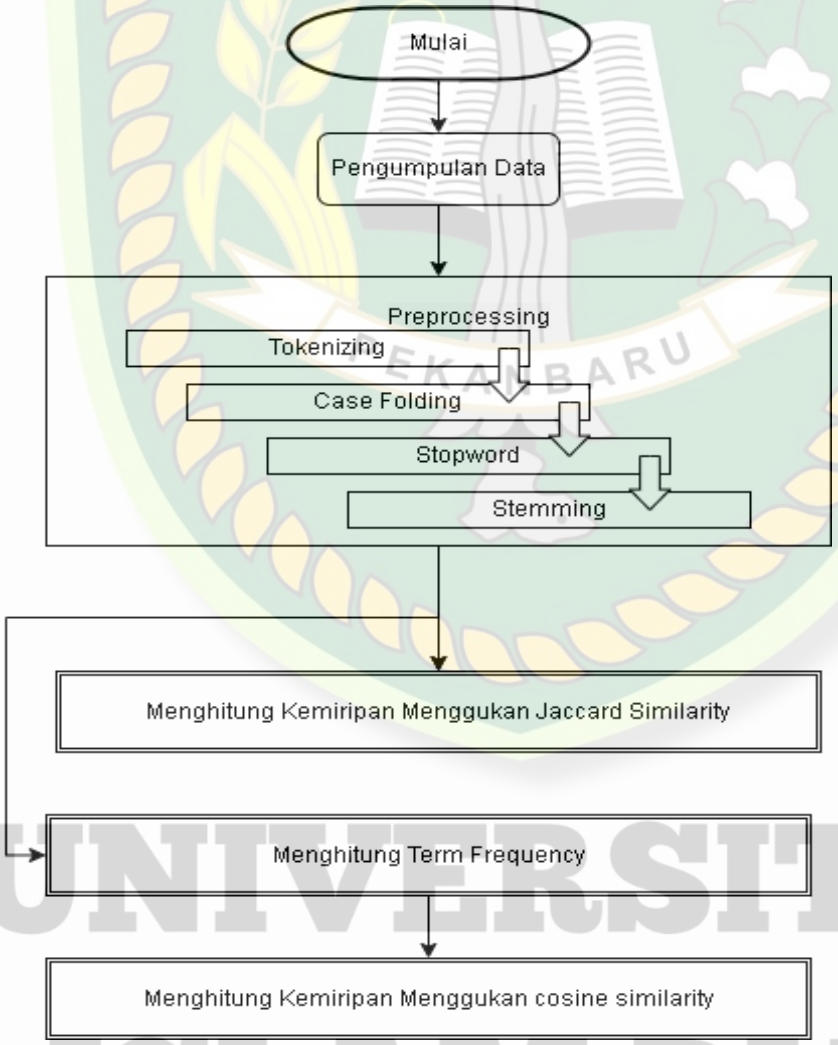
3. Data

Data yang digunakan dalam penelitian ini akan diambil dari berbagai sumber seperti repositori universitas, jurnal ilmiah online, dan artikel-artike

l terkait. Dokumen-dokumen ini mencakup berbagai topik dan akan digunakan untuk membentuk dataset uji. Pengumpulan data akan memastikan keberagaman dan representasi yang sesuai dengan tujuan penelitian.

3.3 Tahapan Penelitian

Tahapan penelitian yang dilakukan melalui beberapa tahapan yaitu pengumpulan dataset, preprocessing, perhitungan nilai kemiripan dengan metode *Jaccard* similarity dan *Cosine* similarity. Alur penelitian dapat dilihat pada gambar 3.1.



Gambar 3.1 Alur Cara Kerja Sistem

3.4. Pengembangan Sistem Pendeteksi Kemiripan

Sistem pendeteksi kemiripan akan dikembangkan dengan menggunakan bahasa pemrograman Python dan berbagai library terkait seperti *Natural Language Tool Kit (NLTK)*. Komponen-komponen utama dalam pengembangan sistem ini meliputi:

- Pengumpulan data: Data yang digunakan dalam penelitian ini akan diambil dari berbagai sumber seperti repositori universitas, jurnal ilmiah online, dan artikel-artikel terkait unutk membentuk dataset uji.
- Pre-processing: Dokumen-dokumen dalam dataset akan melalui serangkaian langkah pre-processing. Langkah-langkah ini termasuk tokenization (memecah teks menjadi token atau kata-kata), case folding (mengubah huruf menjadi huruf kecil), filtering (menghilangkan kata-kata stopword) dan stemming (mengubah kata-kata menjadi bentuk dasarnya).
- Pengukuran kemiripan dokumen teks menggunakan *Jaccard similarity* dan *Cosine similarity.*

3.4.1 Pengumpulan Data

Tahap awal dari penelitian ini adalah dengan mengumpulkan data yang akan dipergunakan untuk dibandingkan, dimana terdapat 2 dokumen yang akan dibandingkan, Sebagai contoh:

- Dokumen 1: Surat Adalah Sarana Menyampaikan Informasi Tertulis Oleh satu Pihak Ke Pihak Yang Lain.
- Dokumen 2 : Dokumen Adalah Sebuah Tulisan Yang Memuat Informasi.

3.4.2 Preprocessing

Tahap preprocessing teks dalam penelitian ini melalui 4 tahapan yaitu, tokenizing, case folding, stopwords, dan stemming.

A. Tokenezing

Pada tahap ini proses pemotongan kalimat menjadi setiap kata yang menyusunnya. Potongan- potongan kata ini disebut token.

Tabel 3. 1 Ilustrasi Tokenizing

Data Input Dokumen 1	Data Input Dokumen 2
Surat Adalah Sarana Menyampaikan Informasi Tertulis Oleh satu Pihak Ke Pihak Yg Lain	Dokumen Adalah Sebuah Tulisan Yang Memuat Informasi

Hasil Tokenizing Dokumen 1	Hasil Tokenizing Dokumen 2
Aurat Adalah Sarana Menyampaikan Informasi Tertulis Oleh Satu Pihak Ke Pihak Yang Lain	Dokumen Adalah Sebuah Tulisan Yang Memuat Informasi

B. ***Case Folding***

Pada tahap ini mengubah semua huruf menjadi huruf kecil. Case folding merupakan proses mengubah semua huruf di dalam dokumen menjadi huruf kecil.

Tabel 3. 2 Ilustrasi Case Folding

Hasil Tokenizing Dokumen 1
Surat
Adalah
Sarana
Menyampaikan
Informasi
Tertulis
Oleh
satu
Pihak
Ke
Pihak
Yang
Lain

Hasil Tokenizing Dokumen 2
Dokumen
Adalah
Sebuah
Tulisan
Yang
Memuat
Informasi

Hasil Case Folding Dokumen 1
surat
adalah
sarana
menyampaikan
informasi
tertulis
oleh
satu
pihak
ke
pihak
yang
lain

Hasil Case Folding Dokumen 2
dokumen
adalah
sebuah
tulisan
yang
memuat
informasi

C. *Stopword*

Pada tahap ini menghilangkan sebuah array kata kunci yang dianggap tidak penting atau kata umum yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna. Seperti yang, dan, di, dari, dan lain-lain.

Tabel 3.3 Ilustrasi Stopword

Hasil Case Folding Dokumen 1	Hasil Case Folding Dokumen 2
surat adalah sarana menyampaikan informasi tertulis oleh satu pihak ke pihak yang lain	dokumen adalah sebuah tulisan yang memuat informasi

Hasil Stopword Dokumen 1	Hasil Stopword Dokumen 2
surat - sarana menyampaikan informasi tertulis - satu pihak - pihak - lain	dokumen - sebuah tulisan - memuat informasi

D. ***Stemming***

Pada tahap ini mengemblikan kata-kata yang diperoleh dari hasil filtering ke bentuk dasarnya, menghilangkan imbuhan awal (prefix) dan imbuhan akhir (sufix) sehingga didapat kata dasar.

Tabel 3. 4 Ilustrasi Stemming

Hasil Stopword Dokumen 1	Hasil Stopword Dokumen 2
surat sarana menyampaikan informasi tertulis suatu pihak pihak lain	dokumen sebuah tulisan memuat informasi

Hasil Stemming Dokumen 1	Hasil Stemming Dokumen 2
surat sarana sampai informasi tulis suatu pihak pihak lain	dokumen buah tulis muat informasi

3.4.3 Pengukuran Kemiripan Menggunkan *Jaccard* Similarity

Pada perhitungan *Jaccard*, nilai |A∩B| merupakan jumlah kata yang sama antara dokumen A dengan dokumen B. Untuk dapat mengetahui nilai dokumen A

sama dengan dokumen B dilakukan penyimpanan setiap nilai pada dokumen A kemudian dibandingkan dengan setiap nilai pada dokumen B, apabila sesuai maka nilai irisan ditambahkan dan disimpan. Bila semua nilai sudah dibandingkan maka proses berhenti dan nilai |A∩B| sudah diketahui. Pada proses |AUB| bisa dihitung dengan mencari nilai jumlah kata pada dokumen A kemudian ditambah dengan jumlah kata pada dokumen B dan dikurangi oleh nilai |A∩B|. Misalkan terdapat dua buah dokumen yang dilakukan pembobotan dan pengindeksan. Isi dari kedua dokumenyang digunakan tersebut dapat dirincikan sebagai berikut:

Dokumen 1 = surat sarana sampai informasi tulis satu pihak pihak lain

Dokumen 2 = dokumen buah tulis muat informasi

Kemudian memisahkan dua dokumen di atas menjadi array; (1)surat, (2)sarana, (3)sampai, (4)informasi ,(5)tulis, (6)satu, (7)pihak, (8)pihak, (9)lain, (10)dokumen, (11)buah, (12)muat, (13)tulis, (14)informasi. Berarti memiliki dua set yang berbeda yaitu Dokumen 1 dan Dokumen 2.

Dokumen1 = A,

Dokumen2 = B

A= {1,2,3,5,6,7,8,9} dan B= {4,5,10,11,12} .

Kemudian mencari Union dari kedua dokumen tersebut. Union adalah jumlah kata secara keseluruhan dari dua dokumen yang sedang dihitung. Dari array diatas bisa lihat bahwa jumlah kata secara keseluruhan adalah 14 kata. Union dari Dokumen 1 dan 2 adalah sebagai berikut :

AUB = {1,2,3,4,5,6,7,8,9,10,11,12}

Keterangan :

U = Union

A = Dokumen 1

B = Dokumen 2

Setelah berhasil mendapatkan hasil Union, selanjutnya adalah mencari Intersection diantara dua dokumen tersebut. Intersection adalah jumlah kata yang sama dari dua dokumen yang sedang dihitung. Jika dilihat dari Dokumen A dan Dokumen B, ada beberapa kata yang sama dari kedua dokumen tersebut, antara lain : [4]tulis, [5]informasi , . Intersection dari Dokumen A dan B adalah :

A ∩ B ={4,5}

Keterangan :

∩ = Intersection

A = Dokumen 1

B = Dokumen 2

Union = 1,2,3,4,5,6,7,8,9,10,11,12,13,14 = 14 kata

Intersection =4,5 = 2 kata

Langkah selanjutnya adalah menghitung kemiripan dari kedua dokumen tersebut dengan rumus sebagai berikut :

$$Jaccard\ (A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{2}{14} = 0{,}14$$

Berdasarkan nilai persamaan yang diperoleh maka dapat ditetapkan bahwa nilai kemiripan dari dokumen 1 dan dokumen 2 adalah 0,14.

3.4.4 Pengukuran Kemiripan Menggunkan *Cosine* Similarity

Pada consine similarity algoritma yang dapat digunakan untuk melakukan perhitunxgan kesamaan dari sebuah dokumen. Untuk notasi himpunan pada consine similarity digunakan rumus *Cosine* similarity dapat dirumuskan sebagai berikut:

$$Similatiry\ (A,B) = = \frac{A.B}{\|A\|.\|B\|} = \frac{\sum_{i=1}^{n} A.B}{\sqrt{\sum_{i=1}^{n}(A)^2}.\sqrt{\sum_{i=1}^{n}(B)^2}}$$

Di mana :

A = Vektor A, yang akan dibandingkan kemiripannya

B = Vektor B, yang akan dibandingkan kemiripannya

A.B = dot product antara vektor A dan vektor B

|A| = panjang vektor A

|B| = panjang vektor B

|A||B| = cross product antara |A| dan |B|

Rumus tersebut digunakan untuk mencari kemiripan antar teks dokumen.

Sebagai contoh:

D1: surat sarana pihak sampai informasi surat tulis satu pihak pihak lain

D2: dokumen buah tulis muat buah informasi

Setelah tahapan pre-processing dilakukan, maka selanjutnya akan diimplementasikan metode TF-IDF terlebih dahulu. Tahapan awal dari metode TF-IDF adalah melakukan proses pembobotan. pembobotan kata weighting diawali dengan menghitung jumblah kata dalam dokumen. Berikut ini Tabel 1 menunjukkan hasil term frequency.

Tabel 3. 5 Pembobotan Kata

Term	D1	D2
Surat	2	0
Sarana	1	0
Sampai	1	0
Informasi	1	1

Term	D1	D2
Tulis	1	1
Satu	1	0
Pihak	3	0
lain	1	0
Dokumen	0	1
Buah	0	2
Muat	0	1

Vektor D1 = 1, 1, 2, 1 ,1 ,1 ,3 ,1 ,0 ,0 ,0

Vektor D2 = 0, 0, 0, 1, 2, 0, 0, 0, 1, 2, 1

$$= \frac{(1*0)+(1*0)+(1*0)+(1*1)+(1*1)+(1*0)+(3*0)+(1*0)+(0*1)+(0*1)+(0*1)}{\sqrt{1^{2+}\,1^{2+}\,2^{2+}\,1^{2+}\,1^{2+}\,1^{2+}\,3^{2+}\,0^{2+}\,0^{2+}\,0^{2+}\,0^{2}} * \sqrt{0^{2+}\,0^{2+}\,0^{2+}\,1^{2+}\,2^{2+}\,0^{2+}\,0^{2+}\,0^{2+}\,1^{2+}\,2^{2+}\,1^{2}}}$$

$$= \frac{0+0+0+1+1+0+0+0+0+0}{\sqrt{18} * \sqrt{11}}$$

$$= \frac{2}{14}$$

$$= \frac{2}{14}$$

=0,14

Berdasarkan nilai persamaan yang diperoleh maka dapat ditetapkan bahwa nilai kemiripan dari dokumen 1 dan dokumen 2 adalah 0,14 atau 14 %.

BAB IV

HASIL DAN PEMBAHASAN

Pada penelitian ini disusun secara terstruktur agar setiap tahapan penilitian mendapatkan hasil yang diharapkan. Adapun data yang didapatkan berasal dari beberapa sumber yang terdapat di internet seperti paper ilmiah,sistem purse uir, tugas akhir, literatur jurnal.

4.1 Hasil Penelitian

Dalam penelitian ini melakukan perbandingan antara algoritma *Cosine* similarity dan *Cosine* similarity untuk mengetahui metode terbaik dalam melakukan pengukuran kemiripan abstrak untuk sisten purse uir, Setelah dilakukan serangkaian ujicoba dengan menggunakan perhitungan kemiripan *Cosine* similarity dan *Jaccard* maka diperoleh hasil sebagai berikut :

4.1.1 Hasil Untuk Percobaan 10 Kata

Dari hasil uji coba menghitung kemiripan dokumen text dari algoritma *Cosine* similarity dan *Jaccard* similarity dari 6 dokumen dengan dengan panjang kata masing-masing dokumen adalah 10 kata . Maka hasil dari pengujian tersebut yaitu :

Tabel 4. 1 Tabel Pengujian 10 Kata

Kode	*Jaccard* similarity	*Cosine* similarity	Selisih
D1	0.225764	0.222222	-0.003542
D2	0.5	0.503102	0.003102
D3	0.3	0.304125	0.004125

D4	0.3	0.304125	0.004125
D5	0.153846	0.159764	0.005918
Rata-Rata			0.002745

4.1.2 Hasil Untuk Percobaan 20 Kata

Dari hasil uji coba menghitung kemiripan dokumen text dari algoritma *Cosine* similarity dan *Jaccard* similarity dari 6 dokumen dengan dengan panjang kata masing-masing dokumen adalah 20 kata . Maka hasil dari pengujian tersebut yaitu :

Tabel 4. 2 Tabel Pengujian 20 Kata

Kode	*Jaccard* similarity	*Cosine* similarity	Selisih
D1	0.2	0.201993	0.001993
D2	0.095238	0.096422	0.001184
D3	0.0	0.0	0
D4	0.105263	0.101002	-0.004261
D5	0.04	0.040609	0.000609
Rata-Rata			-000095

4.1.3 Hasil Untuk Percobaan 50 Kata

Dari hasil uji coba menghitung kemiripan dokumen text dari algoritma *Cosine* similarity dan *Jaccard* similarity dari 6 dokumen dengan dengan panjang kata masing-masing dokumen adalah 50 kata . Maka hasil dari pengujian tersebut yaitu:

Tabel 4. 3 *Tabel Pengujian 50 Kata*

kode	*Jaccard similarity*	*Cosine similarity*	Selisih
D1	0.138888	0.194686	0.055798
D2	0.629629	0.727111	0.097482
D3	0.232558	0.234125	0.001567
D4	0.085106	0.094339	0.009233
D5	0.095238	0.128539	0.003301
Rata-Rata			0.039418

4.1.4 Hasil Untuk Percobaan 100 Kata

Dari hasil uji coba menghitung kemiripan dokumen text dari algoritma *Cosine* similarity dan *Jaccard* similarity dari 6 dokumen dengan dengan panjang kata masing-masing dokumen adalah 100 kata . Maka hasil dari pengujian tersebut yaitu :

Tabel 4. 4 Tabel Pengujian 100 Kata

Kode	*Jaccard* similarity	*Cosine* similarity	Selisih
D1	0.063291	0.082962	0.019671
D2	0.138888	0.173582	0.034694
D3	0.150684	0.195294	0.04461
D4	0.096385	0.138340	0.041955
D5	0.048780	0.031852	0.016928
Rata-Rata			0.031571

4.1.5 Hasil untuk percobaan 250 kata

Dari hasil uji coba menghitung kemiripan dokumen text dari algoritma *Cosine* similarity dan *Jaccard* similarity dari 6 dokumen dengan dengan panjang kata masing-masing dokumen adalah 250 kata. Maka hasil dari pengujian tersebut yaitu :

Tabel 4. 5 Tabel Pengujian 250 Kata

Kode	*Jaccard* similarity	*Cosine* similarity	Selisih
D1	0.253731	0.365797	0.112066
D2	0.1	0.117910	0.01791
D3	0.121019	0.140419	0.0194
D4	0.187050	0.224505	0.037455
D5	0.041916	0.032882	-0.009034
Rata-Rata			0.031156

4.1.6 Hasil Untuk Percobaan 500 Kata

Dari hasil uji coba menghitung kemiripan dokumen text dari algoritma *Cosine* similarity dan *Jaccard* similarity dari 6 dokumen dengan dengan panjang kata masing-masing dokumen adalah 500 kata. Maka hasil dari pengujian tersebut yaitu :

Tabel 4. 6 Tabel Pengujian 500 Kata

Kode	*Jaccard* similarity	*Cosine* similarity	Selisih
D1	0.077253	0.105327	0.028073
D2	0.239436	0.281408	0.041972

D3	0.171052	0.193963	0.022911
D4	0.135658	0.163427	0.027769
D5	0.125461	0.093415	0.032046
Rata-Rata			0.0305542

4.1.7 Pengujian Term Frecuency

Pada tahap ini dilakukan pengujian pengaruh term frekuency terhadap hasil *Cosine* similarity dan *Jaccard* similarity dan di dapat hasil pengujian sebagai berikut:

Tabel 4. 7 Tabel Pengujian Term Frecuncy

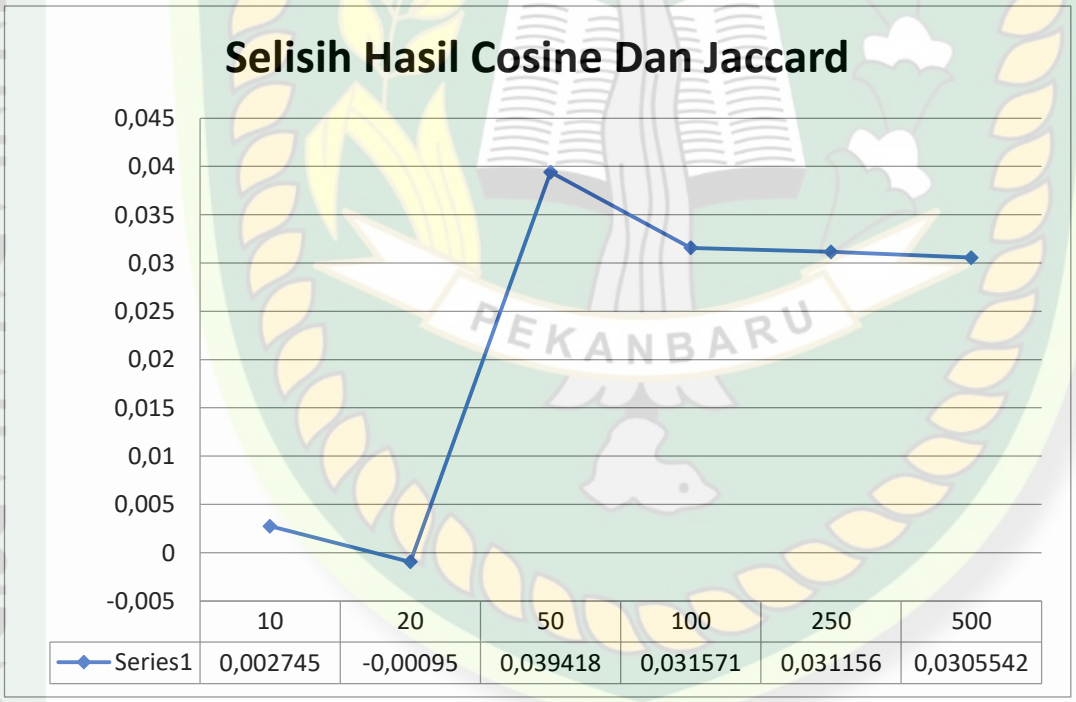
Kode	*Jaccard* similarity	*Cosine* similarity
D1	0.13580246	0.18633581
D2	0.13580246	0.24630694
D3	0.13580246	0.29231484
D4	0.13580246	0.3256774
D5	0.13580246	0.3491579

4.1.8 Grafik Selisih *Cosine* Dan *Jaccard*

Pada gambar 4.1 dapat dilihat grafik selisih dari hasil *Cosine* similarity dan *Jaccard* similarity terhadap pengukuran kesamaaan teks, dimana terdapat perbedaan hasil yang siknifkan pada kata panjang kata kata 50 atau lebih.

Tabel 4. 8 Selisih Hasil

Data	Rata-Rata selisih
10 kata	0.002745
20 kata	-000095
50 kata	0.039418
100 kata	0.031571
250 kata	0.031156
500 kata	0.0305542

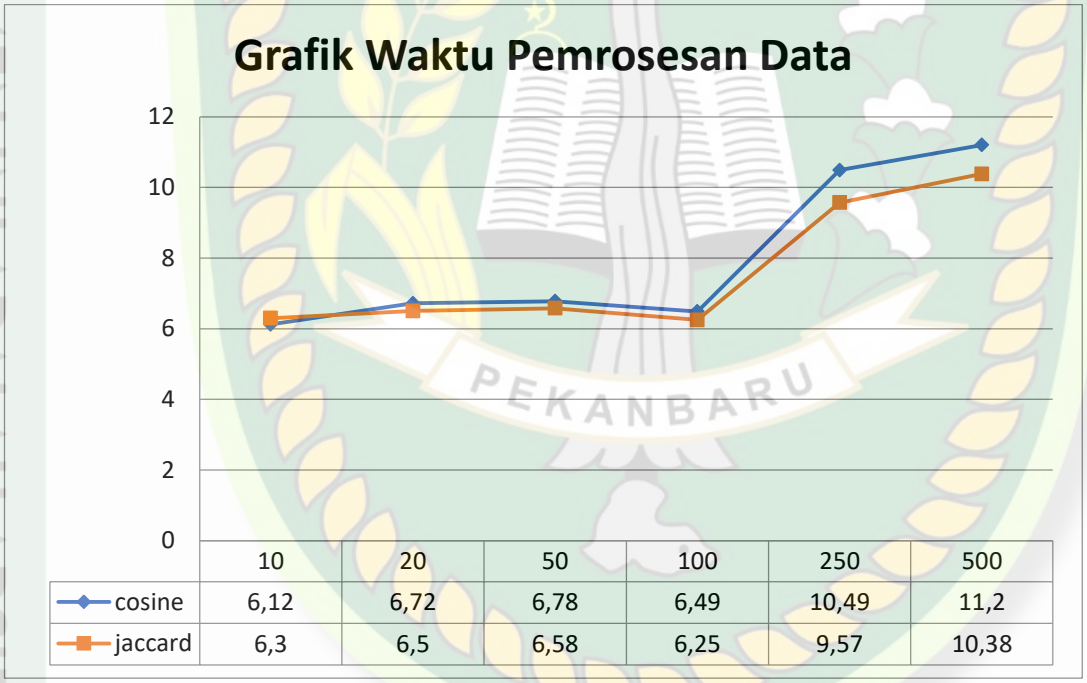


Gambar 4. 1 Grafik Selisih *Cosine* Dan *Jaccard*

4.1.9 Waktu Pemrosesan Data

Hasil pengujian waktu komputasi menggunakan metode *Cosine* similarity dan *Jaccard* simlarity tehadap masing-masing panjang kata dapat dilihat sebagai berikut:

Data	*Jaccard* similarity	*Cosine* similarity
10 kata	6.30s	6.12s
20 kata	6.50s	6.72s
50 kata	6.58s	6.78s
100 kata	6.25s	6.49s
250 kata	9.57s	10.49s
500 kata	10.38s	11.20s



Gambar 4. 2 Gambar grafik waktu pemrosesan data

4.2 Pembahasan

Penelitian ini menggunakan 6 jenis data berdasarkan panjang kata di setiap dokumen. Kemudian dilakukan perbandingan dengan menggunakan algoritma coisne similarity dan *Jaccard* similarity. Implementasi dilakukan pada bahasa pemrograman Python. Adapun implementasi tersebut adalah sebagai berikut:

4.2.1 Dataset

Dataset yang digunakan dalam penelitian ini terdiri dari 36 data yang terbagi menjadi 6 kelompok, yaitu dokumen yang memiliki jumblah 10 kata, 20 kata, 50 kata, 100 kata , 250 kata dan 500 kata. Yang diambil dari berbagai sumber seperti internet, jurnal dan sistem purse UIR.

4.2.2 Preprocessing

Tahap selanjutnya setelah menginput dataset adalah dokumen preprocessing. Preproses dokumen bertujuan membersihkan data dari noise, memiliki dimensi yang lebih kecil, dan lebih terstruktur sehingga dapat diolah lebih lanjut.Tahap preprocessing memiliki beberapa proses yaitu Case Folding, Stop Word Removal, Tokenizing, dan Stemming. Seperti yang dapat dilihat pada gambar 4.1

```
def processing(file):

    #Tokenisasi objek string dari file teks
    tokens = word_tokenize(file)

    #Menghapus tanda baca dan mengubah semua huruf menjadi huruf kecil
    words = [w.lower() for w in tokens if w.isalpha()]

    #Menghapus stopwords dari daftar kata
    stop_words = set(stopwords.words('indonesian'))
    filtered_tokens = [w for w in words if w not in stop_words]

    #Stemming daftar kata
    porter = nltk.PorterStemmer()
    stemmed = [porter.stem(t) for t in filtered_tokens]

    return filtered_tokens


DA = processing(DA)
DB = processing(DB)
DC = processing(DC)
DD = processing(DD)
DE = processing(DE)
DF = processing(DF)
```

Gambar 4. 3 Perintah Untuk Preprocessing Data

4.2.3 Pengukuran Menggunakan *Jaccard* Similarity

Pada perhitungan *Jaccard*, nilai |A∩B| merupakan jumlah fingerprint yang sama antara dokumen A dengan dokumen B. Penghitungan *Jaccard* coeffici ent dilakukan berdasarkan rumus hashing berikut:$J(A, B) = |A \cap B||A \cup B|$, diman a J(A,B) adalah nilai kemiripan antara dataset A dan B, $A \cap B$ adalah irisan/data yang sama dari A dan B, dan $A \cup B$ adalah union/gabungan data dari A dan B. perintah pengukuran menggunakan *Jaccard* similarity dapat diliahat pada gambar 4.2

```
: def jaccard_similarity(list1, list2):
      s1 = set(list1)
      s2 = set(list2)
      return len(s1.intersection(s2)) / len(s1.union(s2))

  print(jaccard_similarity(DA,DB),jaccard_similarity(DA,DC),jaccard_similarity(DA,DD),jaccard_similarity(DA,DE),jaccard_simila
  )
```

Gambar 4. 4 Perintah Pengukuran Menggunkan *Jaccard* Similarity

4.2.4 Pembobotan Term Frequency

Pada tahap ini query pencarian dan dataset artikel ilmiah dilakukan pembobotan kata atau istilah untuk menghitung frekuensi kemunculan setiap kata (term frequency) pencarian pada setiap artikel ilmiah dalam dataset. TF-IDF merupakan metode untuk memberikan bobot setiap term/ kata dasar dengan menghitung frekuensi kemunculan term pada setiap dokumen dalam pencarian informasi cara ini juga dikenal efisien, mudah dan memiliki hasil yang akurat. Penggunaan term frequency dapat dilihat pada gambar 4.3 berikut.

```
: #Mengubah daftar token menjadi string untuk vectorzier
  DA      = ','.join(str(v) for v in DA)
  DB      = ','.join(str(v) for v in DB)
  DC      = ','.join(str(v) for v in DC)
  DD      = ','.join(str(v) for v in DD)
  DE      = ','.join(str(v) for v in DE)
  DF      = ','.join(str(v) for v in DF)
```

Gambar 4. 5 Pembobotan Term Frequency

4.2.5 Pengukuran Menggunakan *Cosine* Similarity

Tahap pengukuran menggunakan *Cosine* Similarity merupakan salah sat u metode pengukuran kesamaan dalam mekanisme sistem temu kembali dokume n. mengukur kesamaan antara dua dokumen atau teks Dalam proses pengukuran kesamaan *Cosine* Similarity menghitung nilai dari sudut yang dihasilkan . Besaran atau nilai yang dihasilkan dari sudut vector antara 0 – 1, dimana semakin mendeka ti 1 maka query dan dokumen memiliki kemiripan yang besar dan semakin mendekati 0 maka memiliki kemiripan yang rendah. *Cosine* similarity dapat pada gambar 4.4 berikut:

```
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics.pairwise import cosine_similarity

doc = [DA,DB]

vectorizer = TfidfVectorizer()
tfidf = vectorizer.fit_transform(doc)
words = vectorizer.get_feature_names_out()
similarity_matrix1 = cosine_similarity(tfidf)

print(similarity_matrix1)
```

Gambar 4. 6 Pengukuran Menggunakan *Cosine* Similarity

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian yang dilakukan dengan cara menguji 6 jenis panjang data menggunakan metode *Cosine similrity* dan *Jaccard similarity*, maka dapat diambil beberapa kesimpulan sebagai berikut:

1. Meskipun terdapat perbedaan hasil antara *Cosine* similarity dan *Jaccard* similarity, rata-rata selisihnya relatif kecil, yaitu sekitar 3%.
2. Selisih hasil antara *Cosine* similarity dan *Jaccard* similarity cenderung stabil pada panjang kata tertentu, dengan fluktuasi sekitar 3%.
3. Untuk waktu pemrosesan data menunjukkan bahwa waktu pemrosesan *Cosine* similarity tetap lebih rendah daripada *Jaccard* similarity, dengan selisih waktu sekitar 1.5 detik pada panjang kata 500. Selisih ini relatif kecil dan dapat diterima.
4. Secara keseluruhan, meskipun terdapat perbedaan hasil antara kedua metode, perbedaan tersebut cenderung kecil dan dapat diterima dan keduanya layak dan dapat di implementasikan dalam sistem purse UIR

5.2 Saran

Berdasarkan hasil penelitian, beberapa saran dapat diajukan untuk penelitian selanjutnya:

1. Penelitian ini dapat diperluas dengan menggali metode- metode lain yang dapat digunakan untuk mengukur kemiripan dokumen, sehingga dapat memperkaya analisa dan pemahaman.

2. Hasil penelitian ini dapat diaplikasikan dalam sistem purse UIR untuk mendeteksi kesamaan abstrak antar proposal penelitian.

Dengan mempertimbangkan saran-saran ini, diharapkan penelitian selanjutnya dapat memberikan kontribusi yang lebih baik dalam pemahaman dan pengembangan metode pengukuran kemiripan dokumen.

DAFTAR PUSTAKA

Abhilash. (2023). *What is the Jaccard Similarity measure in NLP?* iambil kembali dari Educative: https://www.educative.io/answers/what-is-the-*Jaccard*-similarity-measure-in-nlp.

Badruzzaman, M., Mohamad, I., Wildan, B., Wahyudin, D. (2020). Similarity Detection for Hadith of Fiqh of Women using *Cosine* Similarity and Boyer Moore Method. International Journal of Advanced Trends in Computer Science and Engineering, Volume 9, No.1.

Clough, P. (2023). *Plagiarism in natural and programming languages: an overview of current tools and technologies*. Diambil kembali dari Finance:(http://www.dcs.shef.ac.uk/nlp/meter/Documents/reports/plagiarism/Plagiaris m.pdf).

Fajar., A. N, (2020). Penerapan Algoritma *Cosine* Similarity untuk Deteksi Kesamaan Konten pada Sistem Informasi Penelitian dan Pengabdian Kepada Masyaraka. Jurnal Informatika Universitas Pamulang ISSN: 2541-1004 Penerbit: Program Studi Teknik Informatika Universitas Pamulang e-ISSN: 2622-4615 Vol. 5, (529-536).

Ginting, L..(2020). Aplikasi Deteksi Kemiripan Kata Menggunakan Algoritma Rabin- Karp. Jurnal Teknologi dan Informasi (JATI) Volume 12 Nomor 2. Heri Sutikno. 2021. IMPLEMENTASI ALGORITMA *COSINE* SIMILARITY UNTUK MENDETEKSI KEMIRIPAN TOPIK JUDUL. JECSIT, Vol. 1, No. 1,,51-61

Made, S., Putu, J., Kadek, K. (2022). Deteksi Kesamaan Teks Jawaban pada Sistem Test Essay Online dengan Pendekatan Neural Network, Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi) Vol. 5 No. 6.

Nurdiana, O. J. (2019). Perbandingan Metode *Cosine* Similarity Dengan Metode *Jaccard* Similarity Pada Aplikasi Pencarian Terjemah Al-Qur'an Dalam Bahasa Indonesia. *Jurnal Online Informatika*.

Parwita, I. W. (2019). String Matching based Plagiarism Detection for.*International Conference on New*, Media Studies.

Rito, M. (2019). Deteksi Plagiarisme pada Artikel Jurnal Menggunakan Metode *Cosine* Similarity. SMARTICS Journal, Vol.5 No. 1. p22-26.

SaptoU. Dkk. 2022. Deteksi Plagiat Tugas Akhir dengan Metode *Jaccard* Similarity. Jurnal Transistor Elektro dan Informatika (TRANSISTOR EI) Vol. 4.

Rio, P . (2019). Deteksi Plagiarisme pada Artikel Jurnal Menggunakan Meto *Cosine* Similarity. SMARTICS Journal, Vol.5 No. . p22-26

Sunardi1. (.2019). IMPLEMENTASI DETEKSI PLAGIARISME MENGGUNAKAN METODE N-GRAM DAN *JACCARD* SIMILARITY TERHADAP ALGORITMA WINNOWING. TRANSMISI, 20, 2407–6422.

Suhardi, D. (.2020).IMPLEMENTASI DETEKSI PLAGIARISME MENGGUNAKAN METODE *JACCARD* SIMILARITY TERHADAP ALGORITMA WINNOWING. TRANSMISI, 20, 2407–6422.

Samosir, W, V., Fahrul, N. (2023). Rancang Bangun Aplikasi Circle Pertemanan Menururt Minat Dan Hoby Menggunakan Metode *Cosine* Similairity. TEKINFO VOL. 24, NO. 1.

Sapto, U., Imam, M., Andi, R. (2022). Deteksi Plagiat Tugas Akhir dengan Metode *Jaccard* Similarity. Jurnal Transistor Elektro dan Informatika (TRANSISTOR EI), Vol. 4, No. 2.

SURAT KEPUTUSAN DEKAN FAKULTAS TEKNIK UNIVERSITAS ISLAM RIAU
NOMOR : 0606/KPTS/FT-UIR/2023
TENTANG PENGANGKATAN TIM PEMBIMBING PENELITIAN DAN PENYUSUNAN SKRIPSI

DEKAN FAKULTAS TEKNIK

Membaca : Surat Ketua Program Studi Teknik Informatika Nomor : 99/TA-TI/FT/2023 tentang persetujuan dan usulan pengangkatan Tim Pembimbing penelitian dan penyusunan Skripsi.

Menimbang : 1. Bahwa untuk menyelesaikan perkuliahan bagi mahasiswa Fakultas Teknik perlu membuat Skripsi.

2. Untuk itu perlu ditunjuk Tim Pembimbing penelitian dan penyusunan Skripsi yang diangkat dengan Surat Keputusan Dekan.

Mengingat : 1. Undang - Undang Nomor 12 Tahun 2012 Tentang Pendidikan Tinggi
2. Peraturan Presiden Republik Indonesia Nomor 8 Tahun 2012 Tentang Kerangka Kualifikasi Nasional Indonesia
3. Peraturan Pemerintah Republik Indonesia Nomor 37 Tahun 2009 Tentang Dosen
4. Peraturan Pemerintah Republik Indonesia Nomor 66 Tahun 2010 Tentang Pengelolaan dan Penyelenggaraan Pendidikan
5. Peraturan Menteri Pendidikan Nasional Nomor 63 Tahun 2009 Tentang Sistem Penjaminan Mutu Pendidikan
6. Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 49 Tahun 2014 Tentang Standar Nasional Pendidikan Tinggi
7. Statuta Universitas Islam Riau Tahun 2018
8. Peraturan Universitas Islam Riau Nomor 001 Tahun 2018 Tentang Ketentuan Akademik Bidang Pendidikan Universitas Islam Riau

MEMUTUSKAN

Menetapkan : 1. Mengangkat saudara-saudara yang namanya tersebut dibawah ini sebagai Tim Pembimbing Penelitian & penyusunan Skripsi Mahasiswa Fak. Teknik Program Studi Teknik Informatika.

No	Nama	Pangkat	Jabatan
1.	Dr. Arbi Haza Nasution, B.IT (Hons)., M.IT	Lektor	Pembimbing

2. Mahasiswa yang akan dibimbing :

Nama : Nardianas Silalahi
NPM : 193510588
Program Studi : Teknik Informatika
Jenjang Pendidikan : Strata Satu (S1)
Judul Skripsi : Sistem Pendeteksi Kemiripan Dokumen Text Menggunakan Metode Cosine Similarity Dan Jaccard Similarity

3. Keputusan ini mulai berlaku pada tanggal ditetapkannya dengan ketentuan bila terdapat kekeliruan dikemudian hari segera ditinjau kembali.

Ditetapkan di : Pekanbaru
Pada Tanggal : 4 Dzulhijjah 1444 H
23 Juni 2023 M

Dekan,

Dr. Eng. Muslim, ST., MT
NPK : 09 11 02 374

Tembusan disampaikan :
1. Yth. Bapak Rektor UIR di Pekanbaru.
2. Yth. Sdr. Ketua Program Studi Teknik Informatika FT-UIR
3. Arsip

**Surat ini ditandatangani secara elektronik*

F.A.3.10

1

KARTU BIMBINGAN TUGAS AKHIR
SEMESTER GENAP TA 2022/2023

NPM : 193510588

Nama Mahasiswa : NARDIANAS SILALAHI

Dosen Pembimbing : 1. Dr ARBI HAZA NASUTIONB.IT.(Hons), M.IT 2.

Program Studi : TEKNIK INFORMATIKA

Judul Tugas Akhir : SISTEM PENDETEKSI KEMIRIPAN DOKUMEN TEXTMENGGUNAKAN METODE COSINE SIMILARITY DAN JACCARD SIMILARITY

Judul Tugas Akhir (Bahasa Inggris) : TEXT DOCUMENT SIMILARITY DETECTION SYSTEM USING COSINE SIMILARITY AND JACCARD SIMILARITY METHODS

Lembar Ke : ……..

NO	Hari/Tanggal Bimbingan	Materi Bimbingan	Hasil / Saran Bimbingan	Paraf Dosen Pembimbing
1	2 juli 2023	Bab 1,2	Revisi bab 1	
2	2 Agustus 2023	Bab 1 ,3	Revisi Bab 1 Dan 3	
3	16 Agustus	Bab 1,3	Revisi Isentifikasi Masalah Dan Alur Proses Sistembab 3	
4	1 September 2023	Acc Seminar Proposal		
5	14 Desember 2023	Bab 4	Menambahkan chart dan komparasi pada hasil kaccard dan cosine	
6	21 desember 2023	Bab 5	Hasil dan kesimpulan	
7	28 Desember 2023	Acc Sidang Kompre		

Pekanbaru,……..
Wakil Dekan I/Ketua Departemen/Ketua Prodi

MTKZNTEWNTG4

( )

Catatan :

1. Lama bimbingan Tugas Akhir/ Skripsi maksimal 2 semester sejak TMT SK Pembimbingditerbitkan
2. Kartu ini harus dibawa setiap kali berkonsultasi dengan pembimbing dan HARUS dicetak kembali setiap memasuki semester baru melalui SIKAD
3. Saran dan koreksi dari pembimbing harus ditulis dan diparaf oleh pembimbing
4. Setelah skripsi disetujui (ACC) oleh pembimbing, kartu ini harus ditandatangani oleh Wakil Dekan I/ Kepala departemen/Ketua prodi
5. Kartu kendali bimbingan asli yang telah ditandatangani diserahkan kepada Ketua Program Studi dan kopiannya dilampirkan pada skripsi.
6. Jika jumlah pertemuan pada kartu bimbingan tidak cukup dalam satu halaman, kartu bimbingan ini dapat di download kembali melalui SIKAD

SURAT KEPUTUSAN DEKAN FAKULTAS TEKNIK UNIVERSITAS ISLAM RIAU
NOMOR : 0134/KPTS/FT-UIR/2024
TENTANG PENETAPAN DOSEN PENGUJI SKRIPSI MAHASISWA FAK. TEKNIK UNIV. ISLAM RIAU

DEKAN FAKULTAS TEKNIK

Menimbang : 1. Bahwa untuk menyelesaikan studi S.1 bagi mahasiswa Fakultas Teknik Univ. Islam Riau dilaksanakan Ujian Skripsi/Komprehensif sebagai tugas akhir. Untuk itu perlu ditetapkan mahasiswa yang telah memenuhi syarat untuk ujian dimaksud serta dosen penguji.

2. Bahwa penetapan mahasiswa yang memenuhi syarat dan dosen penguji yang bersangkutan perlu ditetapkan dengan Surat Keputusan Dekan.

Mengingat : 1. Undang - Undang Nomor 12 Tahun 2012 Tentang Pendidikan Tinggi
2. Peraturan Presiden Republik Indonesia Nomor 8 Tahun 2012 Tentang Kerangka Kualifikasi Nasional Indonesia
3. Peraturan Pemerintah Republik Indonesia Nomor 37 Tahun 2009 Tentang Dosen
4. Peraturan Pemerintah Republik Indonesia Nomor 66 Tahun 2010 Tentang Pengelolaan dan Penyelenggaraan Pendidikan
5. Peraturan Menteri Pendidikan Nasional Nomor 63 Tahun 2009 Tentang Sistem Penjaminan Mutu Pendidikan
6. Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 49 Tahun 2014 Tentang Standar Nasional Pendidikan Tinggi
7. Statuta Universitas Islam Riau Tahun 2018
8. Peraturan Universitas Islam Riau Nomor 001 Tahun 2018 Tentang Ketentuan Akademik Bidang Pendidikan Universitas Islam Riau

MEMUTUSKAN

Menetapkan : 1. Mahasiswa Fakultas Teknik Universitas Islam Riau yang tersebut namanya dibawah ini :

Nama	: Nardianas Silalahi
NPM	: 193510588
Program Studi	: Teknik Informatika
Jenjang Pendidikan	: Strata Satu (S1)
Judul Skripsi	: Sistem Pendeteksi Kemiripan Dokumen Text Menggunakan Metode Cosine Similarity Dan Jaccard Similarity

2. Penguji Skripsi/Komprehensif mahasiswa tersebut terdiri dari :

1. Dr. Arbi Haza Nasution, B.IT., M.IT.	Sebagai Ketua Merangkap Penguji
2. Ana Yulianti, S.T., M.Kom.	Sebagai Anggota Merangkap Penguji
3. Ause Labellapansa, S.T., M.Cs., M.Kom.	Sebagai Anggota Merangkap Penguji

3. Laporan hasil ujian serta berita acara telah sampai kepada Pimpinan Fakultas selambat-lambatnya 1(satu) bulan setelah ujian dilaksanakan.

4. Keputusan ini mulai berlaku pada tanggal ditetapkannya dengan ketentuan bila terdapat kekeliruan dikemudian hari segera ditinjau kembali.

KUTIPAN : Disampaikan kepada yang bersangkutan untuk dapat dilaksanakan dengan sebaik-baiknya.

Ditetapkan di : Pekanbaru
Pada Tanggal : 21 Rajab 1445 H
01 Februari 2024 M

Dekan,

Prof. Dr. Eng. Ir. Muslim.,ST.,MT.,IPU
NPK : 1016047901

Tembusan disampaikan :
1. Yth. Rektor UIR di Pekanbaru.
2. Yth. Ketua Program Studi Teknik Informatika FT-UIR
3. Yth. Pembimbing dan Penguji Skripsi
3. Mahasiswa yang bersangkutan
5. Arsip

**Surat ini ditandatangani secara elektronik*

YAYASAN LEMBAGA PENDIDIKAN ISLAM (YLPI) RIAU
UNIVERSITAS ISLAM RIAU
FAKULTAS TEKNIK
PROGRAM STUDI TEKNIK INFORMATIKA
Jalan Kaharuddin Nasution No. 113 P. Marpoyan Pekanbaru Riau Indonesia – Kode Pos: 28284
Telp. +62 761 674674 *Website*: www.eng.uir.ac.id *Email*: fakultas_teknik@uir.ac.id

BERITA ACARA UJIAN SKRIPSI

Berdasarkan Surat Keputusan Dekan Fakultas Teknik Universitas Islam Riau, Pekanbaru, tanggal 01 Februari 2024, Nomor: 0134/KPTS/FT-UIR/2024, maka pada hari Rabu, tanggal 31 Januari 2024, telah dilaksanakan Ujian Skripsi Program Studi Teknik Informatika Fakultas Teknik Universitas Islam Riau, Jenjang Studi S1, Tahun Akademik 2023/2024 berikut ini.

1. Nama : Nardianas Silalahi
2. NPM : 193510588
3. Judul Skripsi : Sistem Pendeteksi Kemiripan Dokumen Text Menggunakan Metode Cosine Similarity Dan Jaccard Similarity
4. Waktu Ujian : 08.00 WIB s.d. Selesai
5. Tempat Pelaksanaan Ujian : Ruang Sidang Fakultas Teknik UIR

Dengan keputusan Hasil Ujian Skripsi:

~~Lulus~~* / Lulus dengan Perbaikan* / ~~Tidak Lulus~~*

** Coret yang tidak perlu.*

Nilai Ujian:

Nilai Ujian Angka = 81,20 Nilai Huruf = (A)

Tim Penguji Skripsi.

No	Nama	Jabatan	Tanda Tangan
1	Dr. Arbi Haza Nasution, B.IT., M.IT.	Ketua	1.
2	Ana Yulianti, S.T., M.Kom.	Anggota	2.
3	Ause Labellapansa, S.T., M.Cs., M.Kom.	Anggota	3.

Panitia Ujian
Ketua,

Dr. Arbi Haza Nasution, B.IT., M.IT.
NIDN. 1023048901

Pekanbaru, 01 Februari 2024
Mengetahui,
Dekan Fakultas Teknik

Prof. Dr. Eng. Ir. Muslim, S.T., M.T., IPU.
NIDN. 1016047901

UNIVERSITAS ISLAM RIAU
FAKULTAS TEKNIK
الجامعة الإسلامية الريوية
Alamat: Jalan Kaharuddin Nasution No.113, Marpoyan, Pekanbaru, Riau, Indonesia - 28284
Telp. +62 761 674674 Email: fakultas_teknik@uir.ac.id Website: www.eng.uir.ac.id

SURAT KETERANGAN BEBAS PLAGIAT

Nomor: 035/A-UIR/5-T/2024

Operator Turnitin Fakultas Teknik Universitas Islam Riau menerangkan bahwa Mahasiswa/i dengan identitas berikut:

Nama	:	**NARDIANAS SILALAHI**
NPM	:	193510588
Program Studi	:	Teknik Informatika
Jenjang Pendidikan	:	Strata Satu (S1)
Judul Skripsi TA	:	SISTEM PENDETEKSI KEMIRIPAN DOKUMEN TEXT MENGGUNAKAN METODE COSINE SIMILARITY DAN JACCARD SIMILARITY

Dinyatakan **Bebas Plagiat,** berdasarkan hasil pengecekan pada Turnitin menunjukkan angka **Similarity Index < 30%** sesuai dengan peraturan Universitas Islam Riau yang berlaku.

Demikian surat keterangan ini dibuat untuk dapat dipergunakan sebagaimana mestinya.

Mengetahui,

Kaprodi. Teknik Informatika

Apri Siswanto, M.Kom., Ph.D

Pekanbaru, 24 Januari 2024 M
12 Rojab 1445 H

Operator Turnitin F. Teknik

Ahmad Pandi, S.Kom.