

Integrasi Tf-Idf Dan Algoritma *Cosine Similarity* Untuk Deteksi Tingkat Kemiripan Judul Penelitian (Studi Kasus Mahasiswa Fakultas Ilmu Komputer UNISAN Gorontalo)

Asmaul Husnah Nasrullah
Teknik Komputer, Universitas Negeri Makassar
asmaul.husnah@unm.ac.id

Abstrak: Penjiplakan atau plagiarisme masih sering terjadi, terutama dalam dunia akademik. Penjiplakan banyak dilakukan terutama saat menyelesaikan tugas akhir atau skripsi. Untuk menghindari kegiatan plagiat perlu dilakukan dengan mendeteksi kemiripan atau kesamaan judul TA. Deteksi Similaritas dilakukan untuk memeriksa kesamaan dokumen dengan dokumen lain. Salah satu solusi untuk memeriksa / mendeteksi kemiripan dokumen secara cepat dan tepat adalah dengan penerapan algoritma *Cosine Similarity* dan pembobotan TF-IDF. Penelitian ini bertujuan untuk mengetahui kinerja dari algoritma *Cosine Similarity* dalam mendeteksi tingkat kemiripan tugas akhir atau penelitian. Dengan memanfaatkan preprocessing teks yang terdiri dari case folding, tokenizing, stopword removal, dan stemming yang selanjutnya dilakukan perhitungan pembobotan TF-IDF dan nilai kemiripan menggunakan cosine similarity sehingga mendapatkan nilai persentase kemiripan. Hasil dari percobaan pada penelitian ini adalah berupa tingkat kemiripan dari judul tugas akhir yang dibandingkan, dengan pengujian *confusion matrix* diperoleh nilai *accuracy* memiliki rata-rata sebesar 89.7%, *precision* 72.4%, dan *recall* 94.6%.

Kata Kunci: Deteksi, Kemiripan, Judul, TF-IDF, *Cosine Similarity*.

I. PENDAHULUAN

Perkembangan dunia akademik, khususnya dalam bidang penelitian, semakin pesat seiring dengan kemajuan teknologi informasi dan komunikasi. Salah satu aspek yang menjadi perhatian utama dalam penelitian akademik adalah pemilihan judul penelitian yang relevan, unik, dan belum banyak dilakukan oleh peneliti lain. Dalam konteks ini, deteksi kemiripan judul penelitian menjadi suatu kebutuhan penting untuk memastikan orisinalitas dan kontribusi penelitian yang lebih berarti.

Dalam suatu instansi akademik, penelitian tugas akhir bagi mahasiswa merupakan sks wajib pada setiap program studi diseluruh perguruan tinggi. Tugas akhir ialah suatu hal yang sudah tidak asing lagi dikalangan para mahasiswa semester akhir, karena banyaknya alumni, proses memperoleh judul menjadi lebih sulit. Untuk itu penelitian ini masih berhubungan dengan judul skripsi[1]. Penjiplakan atau plagiarisme sering terjadi terutama dalam dunia akademik, terlebih pada perguruan tinggi atau universitas. Plagiarisme dilakukan untuk mempermudah tugas dengan teknik copy paste tanpa harus memahami atau mendalami materi yang dimaksud, akibatnya banyak plagiarisme yang terjadi terutama saat menyelesaikan tugas akhir atau skripsi[2]. Unsur plagiarisme dalam sebuah penelitian berasal dari judul maupun isi. Bila tidak dideteksi, TA yang bertambah dari tahun ke tahun dapat menyebabkan penelitian yang sudah ada akhirnya diteliti kembali.

Judul TA merupakan bagian penting dalam sebuah penulisan TA, karena berfungsi untuk menggambarkan pokok bahasan dan argumentasi dalam skripsi atau TA[3]. Salah satu cara untuk mengurangi plagiarisme atau penjiplakan adalah dengan pencegahan dan deteksi dini. Pencegahan berarti menjaga dan mencegah agar plagiarisme tidak dilakukan. Hal ini harus dilakukan terutama dalam

sistem pendidikan[2]. Untuk menghindari kegiatan plagiat perlu diketahui apakah judul TA yang dibuat sudah pernah dilakukan sebelumnya dengan mendeteksi kemiripan atau kesamaan judul TA. Deteksi similaritas dilakukan untuk memeriksa kesamaan dokumen dengan dokumen lain[3].

Penelitian tentang kemiripan judul telah banyak dilakukan, agar dapat mengetahui tingkat kemiripan sebuah judul diperlukan metode khusus yang dapat menangani tingkat kemiripan tersebut. Beberapa metode yang sudah pernah digunakan untuk pendeteksian kemiripan ialah *Smith Waterman*, *Single Hierarchical*, *TF-IDF*, dan *Cosine Similarity*.

Dari hasil penelitian yang telah banyak dilakukan, algoritma yang akan digunakan pada penelitian ini adalah *Cosine Similarity* dengan mengkombinasikan metode pembobotan dari TF-IDF. Pembobotan TF-IDF digunakan untuk mencari representasi nilai tiap dokumen dalam koleksi. Sedangkan *Cosine Similarity* diharapkan dapat menangani permasalahan pada penelitian ini. Kelebihan dari algoritma *cosine similarity* adalah mampu mendeteksi tingkat kemiripan, dengan akurasi yang tinggi. Adapun kelemahan algoritma ini yaitu tidak dapat mendeteksi makna (sinonim) dari setiap kata yang dibandingkan[8].

Penelitian ini berfokus pada integrasi kedua metode tersebut untuk mendeteksi tingkat kemiripan judul penelitian mahasiswa di Fakultas Ilmu Komputer Universitas Sanata Dharma (UNISAN) Gorontalo. Dengan menggunakan Tf-Idf untuk representasi teks dan *Cosine Similarity* untuk perhitungan kemiripan, diharapkan dapat memberikan solusi yang efektif dan efisien dalam mengidentifikasi kesamaan judul penelitian, sehingga memudahkan pihak akademik dalam mengelola dan membimbing penelitian mahasiswa.

Penelitian ini bertujuan untuk mengetahui kinerja algoritma *cosine similarity* dalam mendeteksi tingkat

kemiripan judul tugas akhir mahasiswa teknik informatika secara efektif. Sehingga kedepannya dapat mencegah terlebih mengurangi tingkat kemiripan tugas akhir.

II. METODE PENELITIAN

Similarity merupakan suatu hasil analisis terhadap suatu karya ilmiah guna menentukan apakah karya ilmiah tersebut mengandung unsur plagiarisme atau tidak[10]. Terdapat 3 macam teknik penentuan nilai kemiripan dari suatu dokumen antara lain: [8]

1. *Distance-based similarity measure*, adalah pengukuran tingkat kesamaan berdasarkan jarak.
2. *Feature-based similarity measure*, menghitung tingkat kesamaan dalam bentuk feature yang diperbandingkan. *Feature-based Similarity* sering digunakan untuk melakukan pengklasifikasian atau pattern matching pada gambar maupun teks.
3. *Probabilistic-based measure*, penghitungan tingkat kemiripan dengan membandingkan dua set objek dalam bentuk probability. Seperti *Kullback Leibler Distance* dan *Posterior Probability*.

III. HASIL DAN PEMBAHASAN

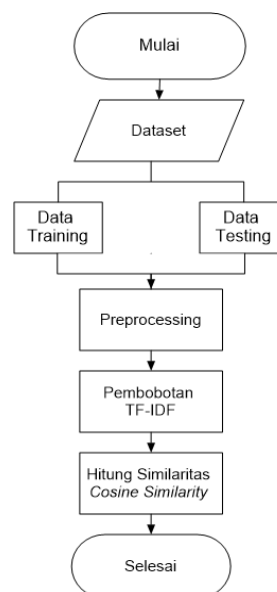
1. Pengumpulan Data

Pengumpulan data merupakan tahap penting dalam proses penelitian ini, yang bertujuan untuk menyediakan informasi yang relevan dan diperlukan dalam analisis kemiripan judul penelitian. Data yang digunakan dalam penelitian ini meliputi judul-judul penelitian mahasiswa Fakultas Ilmu Komputer Universitas Sanata Dharma (UNISAN) Gorontalo yang telah dikumpulkan secara sistematis dan terstruktur.

Data Primer penelitian ini adalah data judul tugas akhir mahasiswa teknik informatika yang didapatkan langsung dari sumbernya yaitu bertempat di Fakultas Ilmu Komputer Universitas Ichsan Gorontalo, yang dikumpulkan dengan teknik observasi dan wawancara. Sedangkan data sekunder yaitu data judul-judul yang ada pada situs google scholar dikumpulkan dengan cara mengambil data-data yang diperoleh dari dokumen, literatur, atau buku-buku yang berhubungan dengan penelitian sebagai pendukung penelitian.

2. Pemodelan

Pemodelan dalam penelitian ini berfokus pada penerapan dua teknik utama dalam Text Mining, yaitu Term Frequency-Inverse Document Frequency (Tf-Idf) untuk representasi teks, dan Cosine Similarity untuk mengukur tingkat kemiripan antar judul penelitian. Kedua teknik ini dipilih karena efektivitasnya dalam mengelola dan menganalisis teks serta menghitung kesamaan antara dua entitas teks secara matematis. Berikut adalah deskripsi rinci mengenai pemodelan yang digunakan dalam penelitian ini.



Gambar 1. Pemodelan *Cosine Similarity*

a. Dataset

Dataset yang digunakan dalam penelitian ini adalah data judul skripsi oleh mahasiswa teknik informatika yang diperoleh dari kantor program studi (prodi) dalam bentuk file excel(.xls). data judul tugas akhir yang digunakan adalah judul tugas akhir 3 tahun terakhir yaitu tahun 2019/2020, 2020/2021, 2021/2022, dan beberapa data judul penelitian yang diperoleh dari situs google scholar dengan tahun yang sama. Terdapat total “352” data judul penelitian mahasiswa. Judul skripsi merupakan data yang akan digunakan sebagai data *training* maupun data *testing* dalam mengukur tingkat kemiripan judul.

b. Preprocessing

Setelah melakukan pengumpulan dataset, tahap selanjutnya yang dilakukan sebelum data diolah, terlebih dahulu dilakukan preprocessing data. Hal ini dilakukan agar data asli berubah menjadi data siap olah untuk tahap selanjutnya. Adapun hal-hal yang dilakukan pada tahap ini yaitu *case folding*, *tokenizing*, *stopword*, dan *stemming*.

c. Pembobotan TF-IDF

Pada tahap ini akan dilakukan perubahan proses perubahan data. Yang tadinya bertipe data string berubah menjadi numerik dengan cara melakukan perhitungan pembobotan untuk mengetahui seberapa sering *term* kata yang muncul dengan menggunakan persamaan (1) dan (2).

d. Cosine Similarity

Hasil daripada pembobotan TF-IDF dimasukkan kedalam persamaan (3) untuk dihitung tingkat kemiripan antara dokumen tes dengan dokumen pembanding menggunakan metode

cosine similarity. Pada tahap ini akan dilakukan pendeteksian kemiripan dengan menggunakan metode *cosine similarity*. Metode *cosine similarity* adalah metode yang digunakan untuk menghitung *similarity* dua buah objek yang didasarkan pada *vector space similarity*[16]. Total data yang akan diolah nantinya menggunakan *tools python*, yang terdiri dari *natural language processing* dan *python sastrawi*. Selain itu juga hasil persentase dari setiap perbandingan dokumen akan digunakan untuk diklasifikasikan kedalam 2 kategori rendah dan tinggi.

e. Evaluasi

Evaluasi dilakukan untuk mengukur kinerja dari metode yang digunakan. Dalam mengukur seberapa baik metode yang digunakan, akan dilakukan dengan menggunakan *confusion matrix* pada seluruh data testing untuk dinilai tingkat akurasinya.

IV. KESIMPULAN DAN SARAN

1. Hasil Pengumpulan Data

Dalam mengumpulkan data, data diambil dari kumpulan judul tugas akhir mahasiswa teknik informatika Universitas Ichsan Gorontalo yang sudah selesai dan beberapa jurnal penelitian teknik informatika dari situs <https://scholar.google.com/>, yang akan dijadikan sebagai bahan uji coba perbandingan judul TA mahasiswa teknik informatika Universitas Ichsan Gorontalo.

2. Tahap Preprocessing

- Case Folding** : *Case Folding* merupakan proses untuk mengkonversi teks ke dalam format huruf kecil (*lowercase*). Hal ini bertujuan untuk memberikan bentuk standar pada teks. Sebagai contoh “Meminjam” menjadi “meminjam”.
- Tokenizing** : *Tokenizing* atau disebut juga tahap *Lexical Analysis* adalah proses pemotongan kalimat teks menjadi bagian-bagian yang lebih kecil, yang disebut token. Sehingga menjadi suatu kata yang berdiri sendiri baik dalam bentuk perulangan maupun tunggal.
- Stopword Removal** : Penghilangan kata-kata tidak deskriptif atau kata *stopword* dari hasil *tokenizing*, untuk mengurangi waktu dan menghindari proses *list* yang panjang guna meningkatkan efisiensi, serta keefektifan dari perolehan informasi.
- Stemming** : *Stemming* adalah proses pengubahan bentuk kata menjadi kata dasar atau tahap mencari *root* kata dari tiap kata hasil *filtering*. Sebagai contoh kata meminjam, dipinjam, meminjamkan, akan distem ke kata dasarnya yaitu “pinjam”.

Tabel 1. Hasil Preprocessing

ID	Hasil Case Folding	Hasil Tokenizing	Hasil Stopword/Filtering	Hasil Stemming
q	identifikasi kematang an buah jambu biji dengan image processing mengguna kan metode knn dan fitur ekstraksi glcm	[identifikasi , [kematang an], [buah] [jambu], [biji], [dengan], [image], [processing] [gunakan , [menggunak an], [metode], [knn], [dan], [fitur], [ekstraksi], [glcm].	identifikasi', 'kematangan', 'buah', 'jambu', 'biji', 'image', 'processing', 'menggunakan', 'metode', 'knn', 'fitur', 'ekstraksi', 'glcm'.	identifik asi, matang, buah, jambu, biji, image, processi ng, guna, metode, knn, fitur, ekstraksi , glcm.
D1	pengenal n kematang an buah pisang mengguna kan metode gray level co- occurrence matrix dan k-nearest neighbor.	[pengenal , [kematang an], [buah], [pisang], [menggunak an], [metode], [gray], [level], [co], [occurrence], [matrix], [dan], [k], [nearest], [neighbor].	pengenaln', 'kematangan', 'buah', 'pisang', 'menggunakan', 'metode', 'gray', 'level', 'co', 'occurrence', 'matrix', 'k', 'nearest', 'neighbor'.	kenal, matang, buah, pisang, guna, metode, gray, level, co, occurrenc e, matrix, k, nearest, neighbor .
..

3. Tahap TF-IDF

Setelah dilakukan tahapan *preprocessing*, selanjutnya dilakukan perhitungan metode dengan menggunakan TF.IDF.

a. Term Frequency

Proses Pertama yang dilakukan ialah menghitung TF (*Term Frequency*) dari setiap dokumen dapat dilihat pada tabel dibawah ini.

Tabel 2. Hasil Perhitungan TF-IDF

Term(t)	Q	D 1	D 2	D 3	D 4	D 5	D 6	D 7	D 8	D 9	D 10
identifi kasi	1	0	0	0	0	0	0	0	0	0	0
image	1	0	0	0	0	0	0	0	0	0	0
..

b. Document Frequency (DF)

Selanjutnya dilakukan tahapan dengan menghitung DF ialah dimana menghitung berapa kali dari setiap kata muncul pada dokumen, proses perhitungannya dapat dilihat pada tabel dibawah ini :

Tabel 3. Hasil Perhitungan Document Frequency

Term(t)	Doc(d)	df
identifikasi	Q	1
Kenal	d1,d3,d5,d8,d9	5
matang	d1,d2,d3,d4	4
..

c. IDF (Invers Document Frequency)

Setelah perhitungan TF.IDF dilakukan dan nilai TF.IDF diketahui, maka dilakukan tahapan selanjutnya ialah perhitungan metode cosine similarity yang dimana cosine similarity merupakan proses dimana untuk mengetahui tingkat kemiripan dokumen.

Tabel 4. Hasil Perhitungan Document Frequency

Term(t)	df	Idf
identifikasi	1	1,041
kenal	5	0,342
..

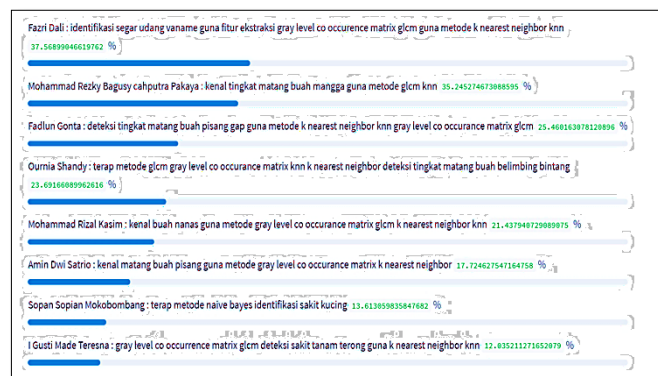
4. Cosine Similarity

Setelah perhitungan TF.IDF dilakukan dan nilai TF.IDF diketahui, maka dilakukan tahapan selanjutnya ialah perhitungan metode cosine similarity yang dimana cosine similarity merupakan proses dimana untuk mengetahui tingkat kemiripan dokumen.

Tabel 5. Hasil Pengurutan Tingkat Kemiripan Dokumen

Dokumen	Nilai Kemiripan
D2	39%
D3	25%
D4	10%
D8	8%
D5	7%
D1	4%
D10	4%
D7	1%
D9	0,08%
D6	0%

Hasil Penerapan Cosine Similarity menggunakan *Sklearn* : Library Machine Learning Untuk Membangun Model pada Python.



Gambar 2. Hasil Penerapan Cosine Similarity pada Python

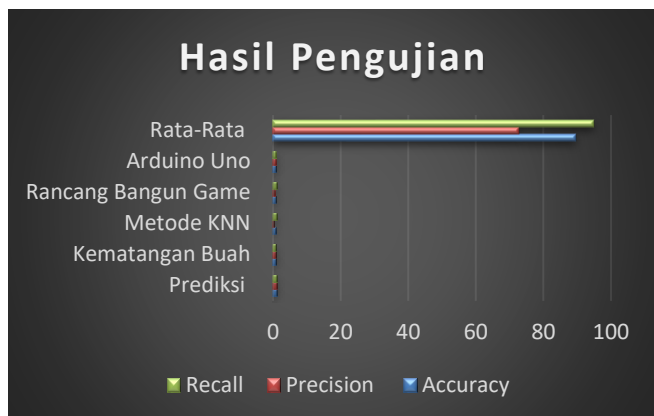
5. Evaluasi Model

Pada Penelitian ini Evaluasi model menggunakan *Accuracy*, *Precision*, dan *Recall*. (*accuracy*) menyangkut perihal pengukuran kinerja dari sebuah metode yang digunakan, (*precision*) berkaitan dengan kemampuan sistem untuk tidak memanggil dokumen yang tidak relevan, sedangkan perolehan (*recall*) berhubungan dengan kemampuan sistem untuk memanggil dokumen yang relevan.

Tabel 6. Hasil Evaluasi Menggunakan Confussion Matrix

	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>
Prediksi	1	1	1
Kematangan Buah	0,94	0,88	0,88
Metode KNN	0,71	0,33	1
Rancang Bangun Game	0,89	0,56	1
Arduino Uno	0,94	0,86	0,86
Rata-Rata	89,7	72,4	94,6

Berdasarkan hasil proses evaluasi yang dilakukan terhadap data *testing* yang berjumlah 5 query pada 35 data dengan nilai *threshold* yang digunakan sebesar 0,3 atau 30%, dapat diperoleh nilai kinerja terhadap model dari algoritma cosine similarity yang digunakan yaitu nilai recall sebesar 94.6%, nilai precision sebesar 72.4%, dan nilai accuracy sebesar 89.7%.



Gambar 3. Hasil Pengujian

Nilai rata-rata hasil pengujian diatas diperoleh dari hasil menjumlahkan tingkat *accuracy* pada masing-masing *query*, kemudian dibagi dengan jumlah *query* yang dijadikan persen. Sehingga dapat disimpulkan berdasarkan perolehan nilai kerja model tersebut menunjukkan bahwa penggunaan algoritma *cosine similarity* dapat bekerja dengan baik dalam mengukur tingkat kemiripan judul tugas akhir, untuk itu penggunaan metode tersebut dapat digunakan dalam membangun aplikasi terhadap bidang yang membutuhkan.

Berdasarkan hasil analisa dan pengujian yang telah dilakukan, maka dapat ditarik kesimpulan bahwa integrasi antara pembobotan TF-IDF dan metode Cosine Similarity dapat menjadi salah satu metode untuk mendeteksi tingkat kemiripan dari judul tugas akhir. Hal ini diperoleh setelah melakukan pengujian dengan membandingkan tiap judul tugas akhir Universitas Ichsan Gorontalo diperoleh rata-rata tingkat kemiripan yang dominan tinggi, dibandingkan dengan perbandingan antara judul artikel jurnal dan judul tugas akhir memiliki nilai tingkat kemiripan rata-rata dominan rendah, hal tersebut dikarenakan penggunaan *stemming* dalam tahap preprocessing yang dapat meningkatkan tingkat kemiripan suatu dokumen.

Pengukuran kinerja/evaluasi dilakukan dengan perhitungan manual, menggunakan *confusion matrix* diperoleh rata-rata akurasi hingga 89.7%. Angka tersebut diperoleh dengan menjumlahkan akurasi pada masing-masing *query* kemudian dibagi dengan jumlah *query* lalu dijadikan persen. Dengan demikian penerapan *Cosine Similarity* dalam mendeteksi kemiripan judul tugas akhir dapat memberikan akurasi yang tergolong dalam kategori baik. Hal ini dibuktikan dengan nilai *accuracy* 89.7%, *precision* 72.4%, dan *recall* 94.6%.

UCAPAN TERIMA KASIH

Puji syukur saya panjatkan kepada Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya, sehingga penelitian ini dapat diselesaikan dengan baik.

Pada kesempatan ini, saya ingin mengucapkan terima kasih yang sebesar-besarnya kepada: Mahasiswa Fakultas Ilmu Komputer UNISAN Gorontalo, yang menjadi objek studi dalam penelitian ini. Terima kasih atas data dan informasi yang telah diberikan untuk kelancaran penelitian ini. Kepada pihak Fakultas Ilmu Komputer UNISAN

Gorontalo, yang telah memberikan kesempatan serta fasilitas yang sangat mendukung untuk pelaksanaan penelitian ini, serta untuk seluruh dosen yang telah memberikan ilmu dan pengalaman berharga selama saya menempuh studi.

Serta ucapan terimakasih kepada semua pihak yang telah memberikan bantuan, baik secara langsung maupun tidak langsung, yang tidak bisa saya sebutkan satu per satu. Semoga kebaikan dan dukungannya selalu mendapatkan balasan yang setimpal.

Semoga hasil penelitian ini dapat memberikan manfaat dan kontribusi yang positif, terutama dalam bidang deteksi kemiripan judul penelitian. Saya menyadari bahwa penelitian ini masih jauh dari sempurna, untuk itu, saya sangat mengharapkan kritik dan saran yang membangun demi perbaikan di masa yang akan datang.

DAFTAR PUSTAKA

- [1] U. N. Hasanah, R. Satra, and F. Umar, "Deteksi Kemiripan Judul Skripsi Menggunakan Algoritma Smith Waterman," *Bul. Sist. Inf. dan ...*, vol. 1, no. 1, pp. 56–65, 2020, [Online]. Available: <http://jurnal.fikom.umi.ac.id/index.php/BUSITI/article/view/676>.
- [2] M. Azmi, "Analisis Tingkat Plagiasi Dokumen Skripsi Dengan Metode Cosine Similarity Dan Pembobotan Tf-Idf," *Tek. Teknol. Inf. dan Multimed.*, vol. 2, no. 2, pp. 90–95, 2022, doi: 10.46764/teknimedia.v2i2.51.
- [3] A. Amrulloh and I. F. Adam, "Sistem Pencarian Similaritas Judul Tugas Akhir Menggunakan Metode TF-IDF," vol. 7, no. 2, pp. 74–82, 2021, doi: 10.24014/coreit.
- [4] R. T. Adek, M. Jannah, J. B. No, and K. B. Indah, "MAHASISWA DENGAN MENGGUNAKAN."
- [5] I. Abdullah and E. Aribowo, "Rancang Bangun Aplikasi Pengecekan Kemiripan Judul Skripsi Dengan Metode Cosine Similarity (Studi Kasus : Program Studi Teknik Informatika UAD)," vol. 6, no. 2, pp. 43–52, 2018.
- [6] N. Andriani and A. Wibowo, "Implementasi Text Mining Klasifikasi Topik Tugas Akhir Mahasiswa Teknik Informatika Menggunakan Pembobotan TF-IDF dan Metode Cosine Similarity Berbasis Web," *Senamika*, no. September, pp. 130–137, 2021, [Online]. Available: <https://conference.upnvj.ac.id/index.php/senamika/article/view/1807%0Ahttps://conference.upnvj.ac.id/index.php/senamika/article/download/1807/1350>.
- [7] F. Septian, "Penerapan Algoritma Cosine Similarity untuk Deteksi Kesamaan Konten pada Sistem Informasi Penelitian dan Pengabd ...," 2016.
- [8] F. B. Sejati, P. Hendradi, and B. Pujiarto, "KEMIRIPAN TEMA MENGGUNAKAN METODE COSINE SIMILARITY (Studi Kasus : Di Universitas Muhammadiyah Magelang)," vol. 2, no. 2, 2019.

- [9] A. Riyani, M. Zidny, and A. Burhanuddin, "Penerapan Cosine Similarity dan Pembobotan TF-IDF untuk Mendeteksi Kemiripan Dokumen," vol. 2, no. 1, pp. 23–27, 2019.
- [10] D. Kurniadi, S. F. C. Haviana, and A. Novianto, "Implementasi Algoritma Cosine Similarity pada sistem arsip dokumen di Universitas Islam Sultan Agung," *J. Transform.*, vol. 17, no. 2, p. 124, 2020, doi: 10.26623/transformatika.v17i2.1613.
- [11] A. Agustiawan, "Analisis Similarity/Kemiripan Artikel Jurnal Online Terbitan Tahun 2019-2020 Di ISI Yogyakarta," *ABDI PUSTAKA J. Perpust. dan Kearsipan*, vol. 2, no. 1, pp. 29–43, 2022, doi: 10.24821/jap.v2i1.6984.
- [12] A. Millah and S. Nurazizah, "Perbandingan Penggunaan Algoritma Cosinus dan Wu Palmer untuk Mencari Kemiripan Kata dalam," vol. 2, no. 1, pp. 15–25, 2017.
- [13] G. E. I. Kambey and Dkk, "Penerapan Clustering pada Aplikasi Pendeteksi Kemiripan Dokumen Teks Bahasa Indonesia," *Penerapan Clust. pada Apl. Pendeteksi Kemiripan Dok. teks Bhs. Indones.*, vol. 15, no. 2, pp. 75–82, 2020.
- [14] B. Herwijayanti, D. E. Ratnawati, and L. Muflikhah, "Klasifikasi Berita Online dengan menggunakan Pembobotan TF-IDF dan Cosine Similarity," vol. 2, no. 1, pp. 306–312, 2018.
- [15] E. L. Amalia, A. J. Jumadi, I. A. Mashudi, and D. W. Wibowo, "Analisis Metode Cosine Similarity Pada Aplikasi Ujian Online Otomatis (Studi Kasus JTI POLINEMA)," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 2, p. 343, 2021, doi: 10.25126/jtiik.2021824356.
- [16] G. E. I. Kambey *et al.*, "Penerapan Clustering pada Aplikasi Pendeteksi Kemiripan Dokumen Teks Bahasa Indonesia," vol. 15, no. 2, pp. 75–82, 2020.