

Implementasi Analisa Kemiripan Teks Untuk Penentuan Dinas Pada Keluhan Warga di Pemerintahan Daerah

Cahya Rahmad¹, M. Hasyim Ratsanjani², Yudistira Eka Putra³
^{1,2,3}Jurusan Teknologi Informasi, Politeknik Negeri Malang, Malang, Indonesia
Korespondensi author: cahya.rahmad@polinema.ac.id

Info Artikel

Diajukan: 1 September 2020
Diterima: 22 September 2020
Diterbitkan: 29 September 2020

Keywords:
Smart System, Cosine Similarity,
Public Complaint

Kata Kunci:
Sistem Cerdas, Kesamaan Cosine,
Pengaduan Masyarakat



Lisensi: cc-by-sa

Copyright © 2020 C. Rahmad, M. H. Ratsanjani, and Y. E. Putra

Abstract

Complaint is submitted by the complainant to the public service complaint manager for implementing services that are not in accordance with service standards, or the disregard of obligations and / or violations of prohibitions by the organizer. To facilitate the management of complaints nationally, the Central Government has launched an online complaints service website called LAPOR and for several other big cities also launched similar websites. Of the 264 million Indonesian population per day there are around 300-570 complaints per day submitted through these websites. This study aims to create a website for citizen complaints using text mining which implements a text similarity analysis for service determination. In this study used text mining methods in the form of text preprocessing and TF-IDF weighting and for analyzing the similarity using the Cosine Similarity method. The results of the study are to produce 75% accuracy by using a similarity approach to complaints with complaints and 80% by using the similarity approach on average for each department.

Abstrak

Pengaduan disampaikan oleh pelapor kepada pengelola pengaduan pelayanan publik atas pelaksanaan pelayanan yang tidak sesuai standar pelayanan, atau pengabaian kewajiban dan/atau pelanggaran larangan yang dilakukan oleh penyelenggara. Untuk memudahkan pengelolaan pengaduan secara nasional, Pemerintah Pusat telah meluncurkan website layanan pengaduan online bernama LAPOR dan untuk beberapa kota besar lainnya juga meluncurkan website serupa. Dari 264 juta penduduk Indonesia per hari, terdapat sekitar 300-570 pengaduan per hari yang disampaikan melalui website tersebut. Penelitian ini bertujuan untuk membuat website pengaduan warga menggunakan text mining yang mengimplementasikan analisis kemiripan teks untuk penentuan layanan. Pada penelitian ini digunakan metode text mining berupa text preprocessing dan pembobotan TF-IDF serta untuk analisis kemiripannya menggunakan metode Cosine Samemility. Hasil penelitian menghasilkan akurasi sebesar 75% dengan menggunakan pendekatan kesamaan pengaduan dengan pengaduan dan 80% dengan menggunakan pendekatan kesamaan rata-rata tiap departemen.

Cara mensitasi artikel:

C. Rahmad, M. H. Ratsanjani, and Y. E. Putra, "Implementasi Analisa Kemiripan Teks Untuk Penentuan Dinas Pada Keluhan Warga di Pemerintahan Daerah," *Jurnal Teknologi Informasi: Teori, Konsep, dan Implementasi (JTI-TKI)*, vol. 11, no. 2, pp. 43-47, October 2020, doi: 10.36382/jti-tki.v11i2.494

PENDAHULUAN

Pengaduan adalah penyampaian keluhan yang disampaikan pengadu kepada pengelola pengaduan pelayanan publik atas pelayanan pelaksana yang tidak sesuai dengan standar pelayanan, atau pengabaian kewajiban dan/atau pelanggaran larangan oleh penyelenggara. Menurut Peraturan Presiden Nomor 76 Tahun 2013 Pasal 1 angka 8 ada 3 (tiga) poin yang bisa diambil yaitu; (1) Penyampaian keluhan tentang pelayanan yang tidak sesuai dengan Standar Pelayanan Publik, (2) Penyampaian keluhan tentang pengabaian kewajiban oleh penyelenggara layanan; (3) Penyampaian keluhan tentang pelanggaran larangan oleh penyelenggara layanan. Dari definisi tersebut dapat kita pahami bahwa pengaduan pelayanan publik yang disampaikan masyarakat kepada penyelenggara pelayanan publik merupakan salah satu

bentuk kontrol atau pengawasan yang dilakukan masyarakat kepada penyelenggara pelayanan publik [1].

Karena begitu pentingnya pengelolaan pengaduan pelayanan publik, hampir diseluruh peraturan yang mengatur tentang pelayanan publik terdapat pasal tentang pengelolaan pengaduan, pemerintah meluncurkan *website* LAPOR sebagai *website* pengaduan keluhan nasional dan beberapa kota besar juga memiliki *website* pengaduan keluhan sendiri seperti untuk kota Malang pada *website* <https://sambat.malangkota.go.id/> [2] dan untuk kota Surabaya pada *website* <http://mediacenter.surabaya.go.id/> [3].

Pada <https://sambat.malangkota.go.id/> ditampilkan data keluhan yang sudah diterima adalah rata-rata jumlah keluhan perharinya sebanyak 300 keluhan dan total keluhan sebanyak 6.207 keluhan sampai tahun ini. Dari data statistik

diatas dapat disimpulkan bahwa data keluhan yang diterima sangatlah tinggi.

Alur proses beberapa *website* tersebut memiliki garis besar yang sama yaitu *Input* (warga menginput pengaduan keluhan), *Terima* (*Admin* menerima pengaduan keluhan dan menverifikasi), *Disposisi* (Pengdisposisian ke dinas yang bersangkutan) dan *Tindak lanjut* (Dinas menindaklanjuti laporan dan menjawab keluhan warga). Dari alur diatas dapat disimpulkan bahwa admin *website* tersebut harus menerima dan mendisposisi setiap keluhan warga yang ada, dibanding dengan jumlah keluhan warga tiap harinya yang mencapai hingga 300—570 keluhan maka diperlukan cara agar proses disposisi bisa dilakukan lebih efisien [4].

Data teks yang diperoleh memerlukan pengolahan awal agar dapat diproses. Data teks akan diproses menjadi data numerik agar dapat dilakukan proses lebih lanjut. Sehingga dalam text mining ada istilah preprocessing data, yaitu proses pendahuluan yang diterapkan terhadap data teks yang bertujuan untuk menghasilkan data numerik. Pada proses *preprocessing* ada beberapa tahapan dimana data keluhan di tangani untuk dapat siap diproses memasuki tahap *text mining* [5]

Beberapa metode dapat digunakan untuk memperoleh kemiripan teks. *Metode Cosine Similarity* merupakan metode yang digunakan untuk menghitung *similarity* (tingkat kesamaan) antar dua buah objek [6]. Secara umum penghitungan metode ini didasarkan pada *vector space similarity measure*. Metode *cosine similarity* ini menghitung *similarity* antara dua buah objek (misalkan D1 dan D2) yang dinyatakan dalam dua buah vektor dengan menggunakan *keywords* (kata kunci) dari sebuah dokumen sebagai ukuran [7]. Salah satu algoritma yang digunakan untuk deteksi kemiripan teks adalah algoritma Ratcliff/Obershelp[8]. Algoritma ini menggabungkan string dari 2 buah teks untuk mendapatkan panjang karakter total (*sequence (string) matching*). Hasilnya digunakan untuk mencari kata yang sama (sub-sequence) dan menghitung panjang karakternya. Sementara itu [9] melakukan penelitian ini dilakukan dengan menerapkan metode Latent Semantic Analysis dalam melakukan pendeteksian Kemiripan Teks. Demikian juga [10] menggunakan metode Algoritma Rabin Karp untuk deteksi kemiripan dokumen. Algoritma Rabin Karp dapat mendeteksi *similarity* (kemiripan) di dalam dokumen. Algoritma ini menggunakan hashing untuk menemukan *substring* (suatu bagian dari *string*) dalam sebuah teks dengan menggunakan K-gram.

Kemiripan satu teks dengan teks yang lain dalam text mining menjadi tantangan untuk dilakukan penelitian. Bagaimana cara mengukurnya dan dengan parameter apa adalah menjadi kajian yang cukup intens. Beberapa teknik telah dilakukan oleh peneliti sebelumnya diantaranya menggunakan pembobotan *Term Frequency-Invers Document Frequency* (TF-IDF)[11]–[14] yaitu perhitungan yang menggambarkan seberapa pentingnya kata (*Term*) dalam sebuah dokumen. Proses ini digunakan untuk menilai bobot relevansi kata dari sebuah dokumen terhadap seluruh dokumen. *Term frequency* adalah ukuran seringnya kemunculan sebuah kata dalam sebuah dokumen. *IDF* (*Inver Document Frequency*) merupakan banyaknya istilah tertentu dalam keseluruhan dokumen [1].

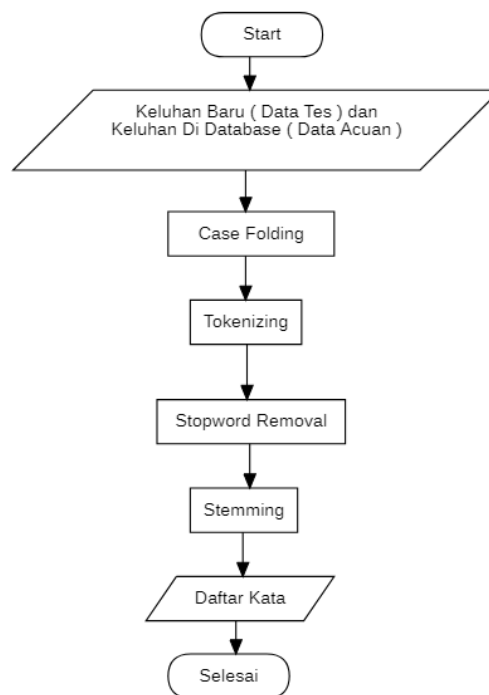
Permasalahan yang dihadapi adalah menghasilkan *website* keluhan pemerintah yang dapat melakukan disposisi keluhan menggunakan text mining dengan mengimplementasikan metode analisis kemiripan, Oleh karena itu tujuannya dibuatnya *website* keluhan pemerintah daerah yang menggunakan implementasi analisa kemiripan teks menggunakan *text mining* dengan metode *Cosine Similarity* dapat membantu proses disposisi keluhan warga pada pemerintahan daerah..

METODE PENELITIAN

Pada penelitian ini penulis menggunakan *Term Frequency-Invers Document Frequency* (TF-IDF) dan *Metode Cosine Similarity* untuk analisa kemiripan teks.

A. Teks mining

Data teks akan diproses menjadi data numerik agar dapat dilakukan proses lebih lanjut. Sehingga dalam *text mining* ada istilah *preprocessing* data, yaitu proses pendahulu yang diterapkan terhadap data teks yang bertujuan untuk menghasilkan data numerik. Pada proses *preprocessing* ada beberapa tahapan dimana data keluhan ditangani untuk dapat siap diproses memasuki tahap *text mining* [5]. Tahap-tahap tersebut adalah



Gambar 1. Flowchart Preprocessing Text

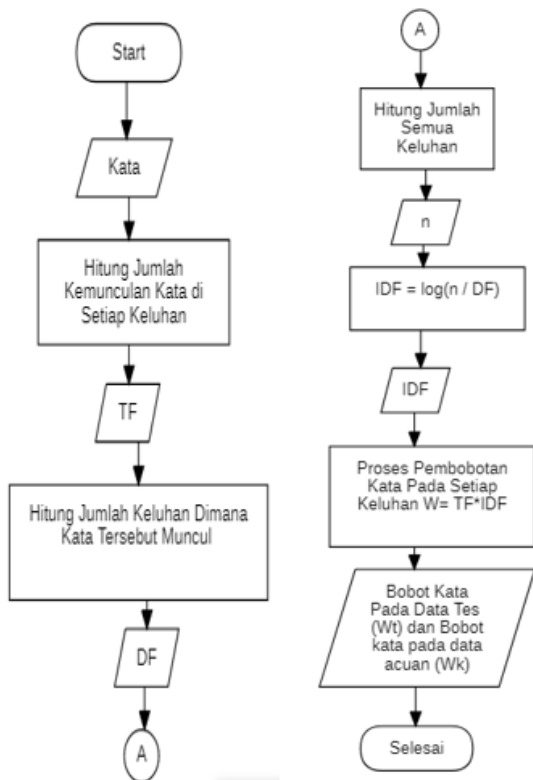
B. *Term Frequency-Invers Document Frequency* (TF-IDF)

TF-IDF yaitu perhitungan yang menggambarkan seberapa pentingnya kata (*Term*) dalam sebuah dokumen. Proses ini digunakan untuk menilai bobot relevansi kata dari sebuah dokumen terhadap seluruh dokumen. *Term frequency* adalah ukuran seringnya kemunculan sebuah kata dalam sebuah dokumen. *IDF* (*Inver Document Frequency*) merupakan banyaknya istilah tertentu dalam keseluruhan dokumen [1], dapat dihitung dengan persamaan 1

$$idf_j = \log \frac{n}{n_j} \quad (1)$$

Keterangan: n = jumlah seluruh dokumen, n_j jumlah dokumen dengan kemunculan istilah j

Gambar 2. Menjelaskan proses pencocokan kata dengan kata yang sama untuk mendapatkan berapa jumlahnya sebagai akumulasi keseringan kemunculan kata dalam keluhan.



Gambar 2 Flowchart Pembobotan TF – IDF

C. Metode Cosine Similarity

Metode *Cosine Similarity*[15] merupakan metode yang digunakan untuk menghitung *similarity* (tingkat kesamaan) antar dua buah objek [6]. Secara umum penghitungan metode ini didasarkan pada *vector space similarity measure*. Metode *cosine similarity* ini menghitung *similarity* antara dua buah objek (misalkan D1 dan D2) yang dinyatakan dalam dua buah vektor dengan menggunakan *keywords* (kata kunci) dari sebuah dokumen sebagai ukuran [7]. Rumus yang digunakan untuk menghitung adalah persamaan 2.

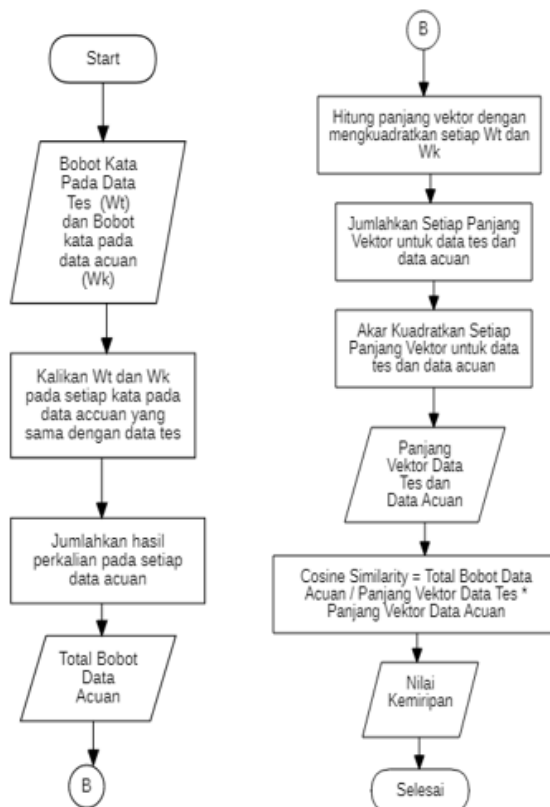
$$\begin{aligned} osSim(d_i, q_i) &= \frac{d_i \cdot q_i}{|q_i| |d_i|} \\ &= \frac{\sum_{j=1}^t (q_{ij} \cdot d_{ij})}{\sqrt{\sum_{j=1}^t (q_{ij})^2 \cdot \sum_{j=1}^t (d_{ij})^2}} \quad (2) \end{aligned}$$

Keterangan :

= bobot istilah j pada dokumen $i = tf_i \cdot idf_j$

= bobot istilah j pada dokumen $i = tf_i \cdot idf_j$

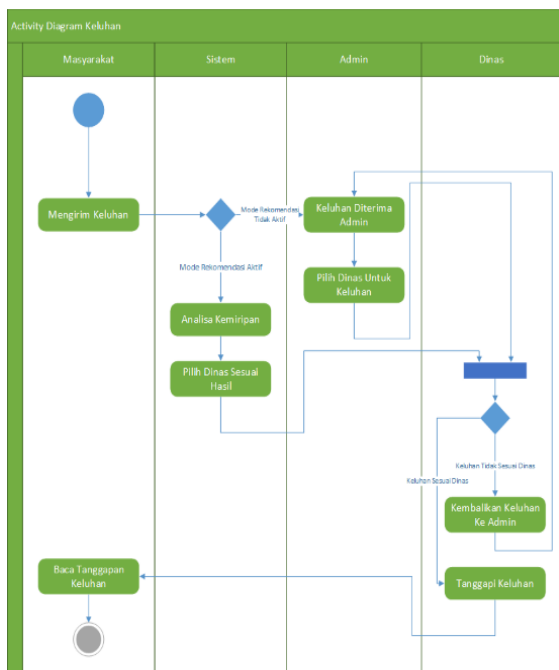
Proses penggunaan metode *Cosine Similarity* pada sistem dijelaskan dengan flowchart dibawah ini



Gambar 3 Flowchart Cosine Similarity

D. Diagram aktifitas

Pada gambar dibawah ini merupakan *activity diagram* yang menjelaskan tentang alur keluhan dalam sistem menurut aktor-aktornya. Mulai dari masyarakat mengirim keluhan yang ingin disampaikan terhadap pemerintah daerah, kemudian dilakukan pengecekan apakah mode rekomendasi aktif atau tidak. Jika mode rekomendasi aktif maka keluhan diproses menggunakan metode analisa kemiripan dan hasil dari analisa kemiripan tersebut akan memilih dinas sesuai dengan hasil yang didapatkan, sedangkan jika mode rekomendasi tidak aktif maka keluhan baru tersebut akan diarahkan ke admin. Keluhan yang sudah masuk ke dinas-dinas akan dicek oleh pegawai dinas terkait, jika keluhan tersebut sesuai maka akan ditanggapi. Tetapi jika terjadi kesalahan dan keluhan tersebut tidak sesuai dengan dinas yang dipilih, maka pegawai dinas dapat mengembalikan keluhan tersebut ke admin. Jika keluhan sudah ditanggapi maka masyarakat bisa melihat hasil tanggapan dari keluhan yang diajukan.



Gambar 4. Diagram aktifitas sistem

E. Mode rekomendasi

Pada penelitian ini mode rekomendasi berfungsi menganalisis kemiripan dan kemudian memilih keluhan yang sesuai dengan jenis dinas [16].

F. Pemilihan dinas

Berdasarkan nilai kemiripan yang didapat antara data tes dan data acuan maka dapat dilakukan pemilihan dinas yang nantinya akan menjadi dinas dari data tes tersebut. Ada dua pendekatan yang dilakukan untuk pemilihan dinas ini:

a. Nilai Kemiripan Keluhan Dengan Keluhan

Pendekatan ini dilakukan dengan cara mencari nilai kemiripan, yang kemudian nilai kemiripan ini dikonversi menjadi bentuk nilai dalam persen. Dari nilai kemiripan ini dicari nilai tertinggi antara data tes dan data acuan. Kemudian dilihat data acuan dengan rangking tertinggi ini berasal dari dinas mana, maka dinas yang akan diberikan ke data tes ini sama dengan dinas dari data acuan tersebut. Contohnya seperti tabel dibawah ini :

Tabel 1 Nilai Kemiripan Keluhan dengan Keluhan

No Keluhan	Isi Keluhan	Nama Dinas	Nilai Kemiripan
1	Akhir akhir sering jadi bakar lihat kok mobil damkar lewat mobil tua beberapa mobil suara sirine dengar pelan kapan reja mobil baru malang banyak gedung tinggi kapan malang punya mobil tangga	Satpol PP	70%
2	selamat pagi maaf mau beri info lampu jalan jalan danau	Dinas Perumahan Dan	60%

sentani raya depan Kawasan
rumah no h1b30 Permukiman
sejak pasang Kota Malang
beberapa tahun lalu
sekarang pernah
nyala mohon
sampai kpd dinas
kait dapat perhati
segera nyala
terimakasih atas
perhati
wassalamualaikum
urus rt01 rw07 kel
madyopuroung
banyak

Jika hasil nilai kemiripan pada data tes seperti tabel diatas maka bisa dilihat nilai kemiripan tertinggi berada pada no keluhan 1 yaitu 70% sehingga data tes akan masuk ke dinas yang sama dengan no keluhan 1 yaitu Satpol PP.

b. Nilai Kemiripan Rata-Rata Setiap Dinas

Pendekatan ini dilakukan dengan cara mencari rata-rata nilai kemiripan yang kemudian nilai kemiripan ini dikonversi menjadi bentuk nilai dalam persen. Dari rata-rata ini kemudian di rangking mana nilai dinas paling tinggi. Contohnya seperti tabel 2 dibawah ini:

Tabel 2 Nilai Kemiripan Rata – Rata Setiap Dinas

No Dinas	Nama Dinas	Rata – rata nilai kemiripan
1	Satpol PP Dinas Perumahan Dan	30%
2	Kawasan Permukiman Kota Malang	60%
3	Dinas Lingkungan Hidup	10%

Pada hasil diatas bisa dilihat bahwa rata-rata nilai kemiripan tertinggi berada pada Dinas Perumahan dan Kawasan Permukiman Kota Malang yaitu 60%. Maka data tes akan diarahkan menuju dinas tersebut.

HASIL DAN PEMBAHASAN

A. Nilai Kemiripan Keluhan dengan Keluhan

Pada penelitian ini menggunakan data tes dan data acuan. Data tes sebanyak 40 keluhan dan data acuan 160 keluhan, dilakukan pengujian pemilihan dinas dengan menggunakan pendekatan nilai kemiripan keluhan dengan keluhan. Hasil dari pengujian pemilihan dinas diatas adalah: 28 keluhan sesuai dinas yang diharapkan, 2 keluhan mempunyai nilai kemiripan yang sama di beberapa dinas, 10 keluhan tidak sesuai dengan dinas yang diharapkan.

B. Nilai Kemiripan Rata – rata Setiap Dinas

Dengan menggunakan data tes dan data acuan diatas yaitu data tes sebanyak 40 keluhan dan data acuan 160 keluhan, dilakukan pengujian pemilihan dinas dengan menggunakan pendekatan nilai kemiripan rata-rata setiap dinas. Pengujian pemilihan dinas ini bisa lihat pada lampiran 5. Hasil dari pengujian pemilihan dinas diatas adalah: 32 keluhan sesuai dinas yang diharapkan, 4 keluhan

mempunyai nilai kemiripan yang sama di beberapa dinas, 4 keluhan tidak sesuai dengan dinas yang diharapkan.

KESIMPULAN

Berdasarkan hasil yang diperoleh dari dari penelitian ini dapat disimpulkan bahwa, Penelitian ini menghasilkan website keluhan pemerintah yang dapat melakukan disposisi keluhan menggunakan text mining dengan mengimplementasikan metode analisis kemiripan. Hasil implementasi metode analisis kemiripan kedalam website yang telah dibuat telah menghasilkan 2 hasil yaitu menggunakan pendekatan nilai kemiripan keluhan dengan keluhan yang mempunyai akurasi 75% dan pendekatan kemiripan rata – rata setiap dinas yang mempunyai akurasi 80%

REFERENSI

- [1.] B. Herwijayanti, D. E. Ratnawati and L. Muflikhah, Klasifikasi Berita Online dengan menggunakan Pembobotan TF-IDF dan, Malang: Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, 2018.
- [2.] P. K. Malang, "Homepage Sambat Pemerintah Kota Malang," 2018. [Online]. Available: <https://sambat.malangkota.go.id/>.
- [3.] P. K. Surabaya, "Homepage Media Center Surabaya," 2018. [Online]. Available: <http://mediacenter.surabaya.go.id/>.
- [4.] S. Rachmatullah and A. P. Wijaya, "Rekomendasi Disposisi Surat dengan Metode Naïve Bayes Pada Arsip," DoubleClick: Journal of Computer and Information Technology, pp. 50 - 59, 2019.
- [5.] S. S. T. Team, "SQL Server Table-valued Functions," 1 Januari 2020. [Online]. Available: <https://www.sqlservertutorial.net/sql-server-user-defined-functions/sql-server-table-valued-functions/>.
- [6.] O. Nurdiana, Jumadi and D. Nursantika, Perbandingan Metode Cosine Similarity dengan Metode Jaccard Similarity Pada Aplikasi Pencarian Terjemah Al-Quran dalam Bahasa Indonesia, Bandung: JOIN | Volume I No. 1 | Juni 2016, 2016.
- [7.] R. T. Wahyuni, D. Prastiyanto and E. Suprpto, Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF pada Sistem Klasifikasi Dokumen Skripsi, Semarang: Jurnal Teknik Elektro Vol. 9 No. 1, 2017.
- [8.] N. Izzah, N. Yusliani, and D. Roodiah, "Sistem Deteksi Kemiripan Teks Pada Berita Berbahasa Indonesia Menggunakan Algoritma Ratcliff / Obershelp," vol. 5, no. 1, pp. 1–6, 2022.
- [9.] B. Ong, D. S. Naga, V. C. M, P. Studi, T. Informatika, and F. T. Informatika, "Kemiripan Teks Dengan Menggunakan," vol. 1, pp. 1–8, 2020.
- [10.] C. Rahmad, P. N. Malang, Y. Ariyanto, and P. N. Malang, "Aplikasi Pendeteksi Kemiripanpada Dokumen Menggunakan," no. May, 2019, doi: 10.33795/jip.v1i2.96.
- [11.] M. A. Ariyanti, A. P. Wibawa, and U. Pujiyanto, "Metode term frequency - invers document frequency pada mekanisme pencarian judul skripsi," *Tekno*, vol. 28, no. 2, p. 177, 2019, doi: 10.17977/um034v28i2p177-190.
- [12.] R. Abidin, E. Sedyono, and S. Y. joko Prasetyo, "Analisis Data Status Di Facebook Untuk Merekomendasikan Produk Makanan Ringan Menggunakan Tf-Idf Dan Document Weighting," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 9, no. 2, pp. 977–984, 2018, doi: 10.24176/simet.v9i2.2518.
- [13.] H. Christian, M. P. Agus, and D. Suhartono, "Single Document Automatic Text Summarization using Term Frequency-Inverse Document Frequency (TF-IDF)," *ComTech Comput. Math. Eng. Appl.*, vol. 7, no. 4, p. 285, 2016, doi: 10.21512/comtech.v7i4.3746.
- [14.] Wiyanto, W. Priatna, and J. S. Hidayat, "Implementasi Term Frequency – Inverse Document Frequency (TF-IDF) Dan Vector Space Model (VSM) Untuk Pencarian Berita Bahasa Indonesia,"

Pelita Teknol. J. Ilm. Inform. Arsit. dan Lingkung., vol. 14, no. November, pp. 119–133, 2019.

- [15] D. Ayu *et al.*, "Pengukuran Kemiripan Dokumen Teks Bahasa Indonesia Menggunakan Metode Cosine Similarity," vol. 9, no. 1, pp. 1–8, 2016.
- [16] I. Sommerville, *Software Engineering Ninth Edition*, 2009.