

## METODE LATENT DIRICHLET ALLOCATION UNTUK MENENTUKAN TOPIK TEKS SUATU BERITA

Sely Karmila , Visda Intan Ardianti

Fakultas Telematika Energi, Institut Teknologi PLN

Email : sely\_karmila@itpln.ac.id

### ABSTRACT

*The development of technology cannot be separated from the news media which is used as a source of public information. Various information contained in news portals both from the region, at home and abroad, as well as various review subjects. News data on online news portals, apart from being a source of information, can also be used as material for analysis and research implementation. Of the various categories discussed in the news portal, there are certain topics, but it is necessary to know a long time to determine which topics are trending in a news portal. This research focuses on the implementation of a news topic system using Latent Dirichlet Allocation. In this study successfully implemented a system where news topics using Latent Dirichlet Allocation resulted in three topics with 67% accuracy results the topics generated were relevant to the document.*

*Keywords: system, modeling topic, news, Latent Dirichlet Allocation.*

### ABSTRAK

Perkembangan teknologi tidak lepas dari media berita yang di jadikan sebagai sumber informasi masyarakat. Berbagai informasi terdapat pada portal berita baik dari daerah, dalam dan luar negeri, serta beraneka ragam subjek ulasan. Data berita pada portal berita online selain sebagai sumber informasi juga dapat digunakan sebagai bahan analisis serta implementasi penelitian. Dari berbagai macam kategori berita yang dibahas dalam portal berita memiliki topik-topik tertentu, namun proses identifikasi memerlukan waktu yang lama untuk menentukan topik mana yang menjadi *trend* dalam suatu portal berita. Penelitian ini berfokus pada implementasi sistem penentuan topik (*modelling topic*) berita menggunakan metode *Latent Dirichlet Allocation*. Dalam penelitian ini telah berhasil menerapkan sistem penentuan topik berita menggunakan *Latent Dirichlet Allocation* menghasilkan tiga topik dengan hasil akurasi 67% topik yang dihasilkan relevan dengan dokumen.

Kata kunci: sistem, *modelling topic*, berita, *Latent Dirichlet Allocation*

## 1. PENDAHULUAN

Berita saat ini di sajikan secara online maupun offline, secara offline dalam bentuk media cetak, lisan, radio, siaran televisi dan secara online melalui media internet seperti portal berita dan lain-lain. Portal berita adalah penyedia media yang menyediakan layanan informasi berupa berita online. Salah satu keuntungan portal berita memiliki adalah ketersediaan informasi yang lebih cepat daripada media tradisional, seperti koran dan majalah.

Berbagai informasi terdapat pada portal berita baik dari daerah, dalam dan luar negeri, serta beraneka ragam subjek ulasan. Data berita pada portal berita online selain sebagai sumber informasi juga dapat digunakan sebagai bahan analisis serta implementasi penelitian. Dari berbagai macam kategori berita yang dibahas dalam portal berita memiliki topik-topik tertentu, namun proses identifikasi memerlukan waktu yang lama untuk menentukan topik mana yang menjadi *trend* dalam suatu portal berita. Oleh karena itu, pada penelitian ini diharapkan dapat memberikan solusi untuk masalah identifikasi topik pada artikel berita yang diterbitkan oleh situs berita online yaitu Detik.com. Topik Pemodelan adalah model statistik yang digunakan untuk menentukan inti dari serangkaian topik dokumentasi.

Pemodelan topik yang akan digunakan dalam penelitian ini adalah *Latent Dirichlet Allocation* (LDA) untuk selanjutnya di implementasikan ke dalam sistem berbasis *web* dengan menggunakan data berita yang bersumber dari portal berita Detik.com. Dataset yang digunakan sebanyak 300

data teks berita dan jumlah kata yang mewakili masing masing topik adalah 30.

## 2. Tinjauan Pustaka

### 2.1. Teks Berita

Teks Berita adalah berita peristiwa yang baru terjadi (Isodarus, 2017). Teks berita terdiri atas tiga bagian, yaitu judul berita, teras berita, dan tubuh berita. Judul berita menginformasikan perihal pokok yang diberitakan. Teras berita menyajikan ringkasan hal-hal yang diberitakan. Tubuh berita menyajikan detail peristiwa yang diberitakan yang menyangkut 5 W + 1 H, what (apa), who (siapa), when (kapan), where (di mana), why (mengapa), dan how (bagaimana).

### 2.2. *Latent Semantic Analysis*

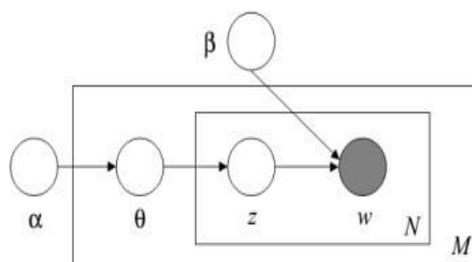
*Latent Semantic Analysis* (LSA) merupakan salah satu algoritma yang sering digunakan dalam proses text mining (Alief, 2018). Menurut Danang Wahyu Wicaksono, dkk. (2014), dijelaskan bahwa *Latent Semantic Analysis* sebagai berikut: LSA adalah suatu metode untuk menemukan hubungan, keterkaitan, dan kemiripan antar dokumen-dokumen, penggalan dari dokumen-dokumen, dan kata-kata yang muncul pada dokumen-dokumen dengan memanfaatkan komputasi statistik untuk menggali dan merepresentasikan konteks yang digunakan sebagai sebuah arti kata untuk sejumlah corpus yang besar. Corpus adalah kumpulan teks yang memiliki kesamaan subyek atau tema.

### 2.3. *Latent Dirichlet Allocation*

*Latent Dirichlet Allocation* merupakan sebuah metode statistika yang digunakan

sebagai model untuk menganalisis suatu dokumen (Faizun Nuril Hikmah, 2020). LDA berusaha untuk melihat dokumen dengan cara mundur untuk menemukan satu set topik yang mungkin telah dikoleksi. LDA merepresentasikan dokumen dengan berbagai topik yang dibuat berdasarkan probabilitas tertentu. Probabilitas topik, merepresentasikan kejelasan dari suatu dokumen.

Representasi model LDA jika digambarkan dalam sebuah diagram dapat digambarkan sebagai berikut:



(Sumber : (Zulhanif et al., 2017))

**Gambar 2.1.** Representasi Model LDA

#### 2.4. Text Mining

*Text Mining* sebagai Sistem Pendeteksi Pemberitaan Palsu yang Tersebar dalam Twitter. Text mining merupakan variasi dari data mining yang berusaha menemukan pola yang menarik dari sekumpulan data tekstual yang berjumlah besar. *Text mining* juga digunakan untuk menangani masalah klasifikasi, clustering, information extraction, dan information retrieval (Tanjung, 2018).

#### 2.5. Preprocessing Text

*Preprocessing text* adalah proses awal terhadap data berupa teks guna menyiapkan data teks tersebut agar data siap untuk diolah

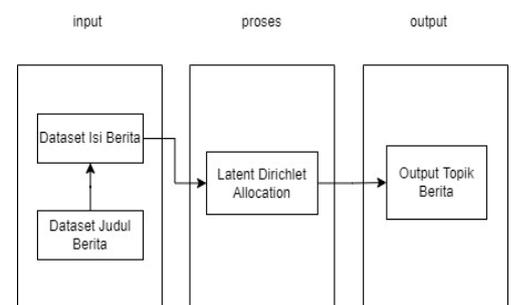
ke tahap selanjutnya (Melita, Amrizal, Suseno, & Dirjam, 2018). Data teks tidak dapat dilakukan klasifikasi secara langsung, sehingga membutuhkan preprocessing text yang bertujuan untuk merubah data teks menjadi data numerik yang siap diolah oleh mesin.

Tahap pembersihan dokumen adalah sebagai berikut:

1. *Tokenizing*
2. *Stopword atau filtering*
3. *Stemming*

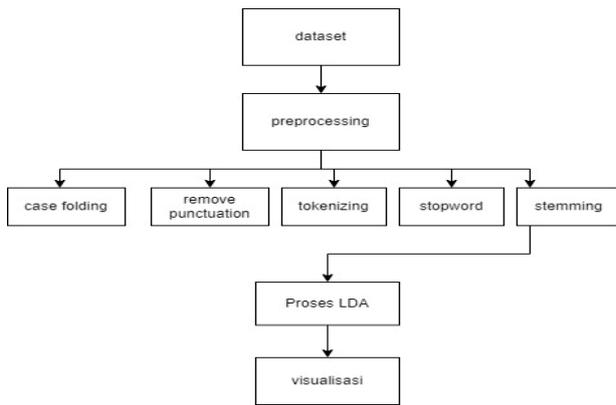
### 3. Perancangan

Adapun desain penelitian yang diterapkan pada penelitian ini adalah sebagai berikut:



**Gambar 3.1** Desain Penelitian

Pada Gambar 3.1. merupakan gambaran desain penelitian, data dimulai dari *input*, yaitu menggunakan dataset isi berita dan dataset judul berita. Kemudian data tersebut akan diproses menggunakan *Latent Dirichlet Allocation*. Setelah diproses maka akan dihasilkan topik teks sebuah berita.



Gambar 3.2 Latent Dirichlet Allocation

Pada gambar 3.2 proses *Latent Dirichlet Allocation* di mulai dari input dataset, selanjutnya data akan dilakukan proses preprocessing meliputi *case folding*, *tokenizing*, *stopword*, serta *stemming*. *Case folding* merubah data teks ke dalam huruf kecil semua supaya data lebih konsisten. Setelah tahap *case folding* dilakukan *remove punctuation* untuk menghilangkan karakter yang tidak terpakai. Selanjutnya data dilakukan proses tokenisasi untuk memisahkan kalimat menjadi kata-kata. *Stopword* digunakan untuk menghilangkan kata-kata tidak penting yang sering muncul seperti “yang”, “dengan”, “di”, dan lain sebagainya. Selanjutnya *stemming* digunakan untuk membentuk kata dasar. Setelah proses tokenisasi selanjutnya pengolahan data dengan menggunakan metode *Latent Dirichlet Allocation (LDA)*. *Output* yang dihasilkan di visualisasikan dengan menampilkan kata-kata yang dihasilkan oleh kelompok-kelompok topik. Visualisasi dengan menggunakan grafik.

4. Hasil

4.1. Pengumpulan Data

Pengumpulan data pada penelitian ini dilakukan dengan mengambil data judul berita serta teks berita yang bersumber dari Detik.com. Dataset berjumlah 300 data dari berbagai kategori berita,

data berupa file excel dengan format xlsx. Tabel 4.1 berikut merupakan sampel data pada penelitian ini:

Tabel 4.1 Sampel Data Penelitian

judul	berita
Produk UMKM Ini Bikin Sandiaga Takjub	Jakarta - Menteri Pariwisata dan Ekonomi Kreatif (Menparekraf) Sandiaga Uno kagum dengan Nurhayati (52), salah satu pelaku UMKM penghasil produk kriya. Pasalnya, Nurhayati mampu mengaik kaum dhuafa untuk membuat kriya yang memiliki nilai jual.Bermula ketika Nurhayati menjelaskan tentang produk kriya berupa lukisan bakar kayu kepada Sandiaga Uno saat workshop Kabupaten Kota (KaTa) di Kota Tangerang Selatan, Banten.Di situ, Nurhayati menjelaskan pembuatan produk kriya bersama kaum dhuafa dengan alat yang terbatas. Nurhayati mengaku kerap meminjam alat ukir berupa pulpen listrik ke rekannya
Ernest Prakasa Sindir Formula E, TGUPP: Ada Iklan Tiket Terjual Habis	Jakarta - Komika Ernest Prakasa mengomentari ajang balap Formula E karena tidak memiliki penggemar. Anggota Tim Gubernur untuk Pembangunan dan Percepatan (TGUPP) DKI Jakarta Tatak Ujiyati merespons kritik yang dilontarkan Ernest lewat Twitter."Tak perlu dikomentari Mas Ernest, toh sudah menghapus cuitannya. Mungkin sudah tahu salah, ya," kata Tatak saat dihubungi, Minggu (5/6/2022). Tatak lalu mengatakan apa yang selama ini dikatakan Ernest terbukti tidak benar.

4.2. Hasil Preprocessing Data

Preprocessing data pada penelitian penulis menggunakan library python untuk proses pembersihan data. Preprocessing data terdiri dari casefolding, stopword, stemming, tokenizing. Berikut merupakan hasil dari preprocessing.

Tabel 4.2. Hasil akhir Preprocessing Data

no	textdata	hasil stemming
1	jakarta menteri pariwisata ekonomi kreatif menparekraf sandiaga uno kagum nurhayati pelaku umkm penghasil produk kriya pasalnya nurhayati mampu	'jakarta', 'menteri', 'pariwisata', 'ekonomi', 'kreatif', 'menparekraf', 'sandiaga', 'uno', 'kagum', 'nurhayati', 'pelaku', 'umkm', 'hasil', 'produk', 'kriya', 'nurhayati'

	mengajak kaum dhuafa membuat kriya memiliki nilai jual bermula nurhayati menjelaskan produk kriya lukisan bakar kayu sandiaga uno workshop	'mampu', 'ajak', 'kaum', 'dhuafa', 'buat', 'kriya', 'milik', 'nilai', 'jual', 'mula', 'nurhayati', 'jelas', 'produk', 'kriya', 'lukis', 'bakar', 'kayu', 'sandiaga', 'uno', 'workshop'
2	jakarta komika ernest prakasa mengomentari ajang balap formula e penggemar anggota tim gubernur pembangunan percepatan tgupp dki jakarta tatak ujiyati merespons kritik dilontarkan ernest lewat twitter perlu dikomentari mas ernest menghapus cuitannya sudah tahu salah kata tatak dihubungi mbak puan terima kasih sempatkan hadir formula e tatak mengatakan dikatakan ernest terbukti benar	'jakarta', 'komika', 'ernest', 'prakasa', 'komentar', 'ajang', 'balap', 'formula', 'e', 'gemar', 'anggota', 'tim', 'gubernur', 'bangunan', 'cepat', 'tgupp', 'dki', 'jakarta', 'tatak', 'ujiyati', 'respon', 'kritik', 'lontar', 'ernest', 'lewat', 'twitter', 'komentar', 'mas', 'ernest', 'hapus', 'cuit', 'tahu', 'salah', 'kata', 'tatak', 'hubung', 'minggu', 'baca', 'anies', 'mbak', 'puan', 'terima', 'kasih', 'sempat', 'hadir', 'formula', 'e', 'tatak', 'kata', 'kata', 'ernest', 'bukti', 'benar'

### 4.3. Hasil Pengolahan Data

Pengolahan data pada penelitian ini dilakukan setelah *preprocessing*, menggunakan metode *latent dirichlet allocation*, metode tersebut untuk menentukan kata topik yang mewakili masing-masing topik. Implementasi metode LDA dilakukan dengan menggunakan library gensim python.

Implementasi LDA menggunakan library gensim dengan membuat objek LDA, menentukan jumlah topik yaitu 3 dan number words adalah jumlah kata yang akan mewakili masing-masing topik yaitu 10. Angka tersebut di ambil karena 3 topik bisa menggambarkan topik yang sedang menjadi *trend* pada portal berita. Untuk menampilkan kata yang dihasilkan per topik digunakan *collection* dan *counter* untuk menghitung jumlah kata per topik.

Berdasarkan pengelompokan kata pada masing-masing topik. Berikut merupakan hasil permodelan topik 1, 2 dan 3

Tabel 4.3. Hasil Modelling Topik

Topik	Hasil Topik Modelling	
Topik 1	0.015*balap + 0.015*motogp + 0.009*mario + 0.009*moto + 0.008*gp + 0.007*motor +  0.006*rossi + 0.006*yamaha +  0.006*honda + 0.005*dunia	balap motogp mario moto gp motor rossi yamaha honda dunia
Topik 2	0.016*mandalika + 0.0014*motogp + 0.013*balap +	mandalika motogp balap hujan helm

	0.010*hujan + 0.010*helm + 0.008*jakarta + 0.008*moto + 0.007*sirkuit + 0.007*formula + 0.006*pawang	jakarta moto sirkuit formula pawang
Topik 3	0.017*mandalika + 0.014*motogp + +0.013*sirkuit + 0.010*balap + 0.010*jokowi+ 0.008*motor + 0.007*gp + 0.007*moto + 0.007*bangun + 0.007*gelar	mandalika motogp sirkuit balap jokowi motor gp moto bangun gelar

Setelah mengetahui hasil modeling topik, selanjutnya dilakukan percobaan untuk mengetahui dominan topik pada setiap dokumen teks, berikut merupakan hasil dominan topik pada 10 dokumen pertama yang di tampilkan pada tabel 4.4

Tabel 4.4. Dominan Topik

No Dokumen	Dominan Topik	Probabilitas
1	3	0.9942
2	2	0.8192
3	2	0.9933
4	2	0.998
5	1	0.9956
6	2	0.9961
7	3	0.7791

8	3	0.9967
9	2	0.9971
10	2	0.9972

Tabel 4.5. Validasi Judul dengan Topik

no	Judul artikel	topik	validasi
1	Produk UMKM Ini Bikin Sandiaga Takjub	mandalika motogp sirkuit balap jokowi motor gp moto bangun gelar	no
2	Ernest Prakasa Sindir Formula E, TGUPP: Ada Iklan-Tiket Terjual Habis	mandalika motogp balap hujan helm jakarta moto sirkuit formula pawang	yes
3	Ernest Prakasa Diserang Netizen Gegara Komentari Formula E	mandalika motogp balap hujan helm jakarta moto sirkuit formula pawang	yes
4	Kualifikasi MotoGP Catalunya: Ukir Rekor Lagi! Aleix Espargaro Pole	mandalika motogp balap hujan helm jakarta moto sirkuit formula pawang	yes
5	Formula E Jakarta Usai, Zita PAN Sedih Tak Lihat Sponsor BUMN	balap motogp mario moto gp motor rossi yamaha honda dunia	no

Hasil pengujian validasi keseluruhan dataset menghasilkan data yang tidak sesuai sebanyak 99 data

tidak sesuai dari jumlah 300 data uji, sehingga menghasilkan akurasi yaitu

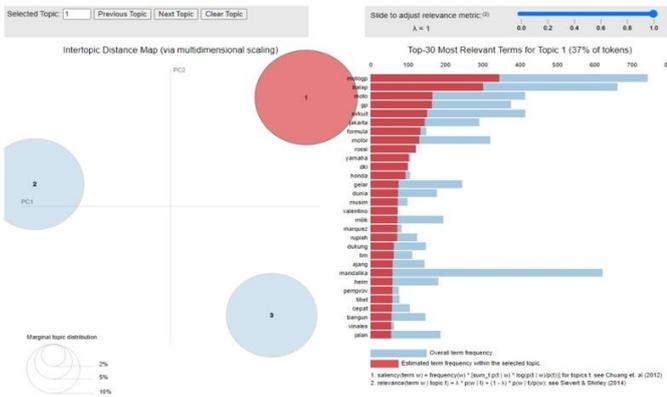
$$akurasi = \frac{total\ data\ sesuai}{total\ data} \times 100\%$$

$$akurasi = \frac{201}{300} \times 100\%$$

$$akurasi = 67\%$$

Dengan demikian tingkat akurasi relevansi topik sebesar 67% dengan tingkat error rate sebesar 33%.

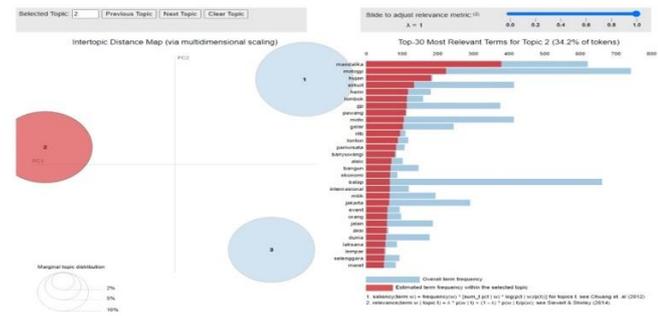
Berikut merupakan tampilan output topik 1 pada gambar 4.1



Gambar 4.1. Halaman Output Topik 1

Berdasarkan gambar 4.1 output topik 1 membahas mengenai balap motogp sirkuit mandalika, valentino rossi, serta dki jakarta. Pada gambar sisi kiri pc1 merupakan sumbu x, dan pc2 sumbu y. Lingkaran angka 1 berwarna merah mewakili topik 1. Pada grafik sisi kanan terdapat angka 37% of tokens yang berarti dari semua dokumen yang mewakili topik 1 sebesar 37% yang relevan. Pada grafik warna merah merupakan jumlah kemunculan kata pada topik 1, serta warna biru merupakan jumlah kemunculan kata pada seluruh dokumen.

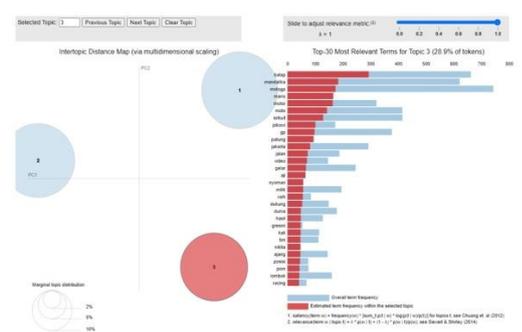
Berikut merupakan tampilan output topik 2 pada gambar 4.2



Gambar 4.2 Halaman output topik 2

Berdasarkan gambar 4.2 output topik 2 membahas mengenai sirkuit mandalika, hujan serta pawang hujan. Pada gambar sisi kiri pc1 merupakan sumbu x, dan pc2 sumbu y. Lingkaran angka 2 berwarna merah mewakili topik 2. Pada grafik sisi kanan terdapat angka 34.2% of tokens yang berarti dari semua dokumen yang mewakili topik 2 sebesar 34.2% yang relevan. Pada grafik warna merah merupakan jumlah kemunculan kata pada topik 2, serta warna biru merupakan jumlah kemunculan kata pada seluruh dokumen.

Berikut merupakan tampilan output topik 3 pada gambar 4.3



Gambar 4.3 Halaman Output Topik 3

Berdasarkan gambar 4.3 output topik 3 membahas mengenai ajang balap mandalika, jokowi, jakarta. Pada gambar sisi kiri pc1 merupakan sumbu x, dan pc2 sumbu y. Lingkaran angka 3 berwarna merah mewakili topik 3. Pada grafik sisi kanan terdapat angka 28.9 % of

tokens yang berarti dari semua dokumen yang mewakili topik 3 sebesar 28.9% yang relevan. Pada grafik warna merah merupakan jumlah kemunculan kata pada topik 3, serta warna biru merupakan jumlah kemunculan kata pada seluruh dokumen.

## 5. Kesimpulan dan Saran

Berikut merupakan kesimpulan penelitian:

Penelitian ini telah berhasil mengimplementasikan metode latent Dirichlet allocation untuk memodelkan topik berita yang bersumber dari detik.com dengan berbagai macam kategori. Yang menghasilkan 3 model topik yang Hasil pemodelan divisualisasikan dalam bentuk grafik untuk mengetahui trend topik yang sedang menjadi pemberitaan dengan hasil pengujian relevansi judul dengan topik menghasilkan akurasi 67%.

## Saran

Berikut merupakan saran dari penelitian guna pengembangan penelitian yang akan datang: Penelitian menggunakan sumber berita online yang lain atau dari berbagai artikel berita yang berbeda guna melihat variasi trend topik yang sedang di bahas.

## DAFTAR PUSTAKA

- Alief, M. S. (2018). Sistem Bantu Deteksi Plagiarisme Judul Skripsi Menggunakan Latent Semantic Analysis.
- Ardi Isbad Amar Gurning, Z. A. (2016). Penerapan Fuzzy String Matching Pada Aplikasi Pencarian Tugas Akhir Mahasiswa Jurusan Sistem Informasi Berbasis Web (Studi Kasus: Fakultas Sains dan Teknologi UIN Suska Riau). *Jurnal Rekayasa dan Manajemen Sistem Informasi*, 54-59.
- Buldan, A. (2021). *Aplikasi Trend Berita Bisnis Berbasis Pemilihan Topik Menggunakan Algoritma Latent Dirichlet Allocation*. Jakarta: Tugas Akhir : Universitas Islam Negeri Syarif Hidayatulloh.
- ER, N. A. (2021). Pemodelan Topik dengan Latent Dirichlet Allocation (LDA) untuk Klasterisasi Cerita Berbahasa Bali. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, 1-8.
- Faizun Nuril Hikmah, S. B. (2020). Deteksi Topik Tentang Tokoh Publik Politik Menggunakan Latent Dirichlet Allocation. *REPOSITOR*, 1-12.
- Grinberg, M. (2014). *Flask Web Development\_ Developing Web Applications*.
- Hanif, I. (2019, April 17). *Fuzzy String Matching: Metode Pencocokan String Berbasis Levenshtein Distance*. Retrieved from medium.com: <https://medium.com/@iqbalhannif/fuzzy-string-matching-metode-pencocokan-string-berbasis-levenshtein-distance-4ba1293871d8>
- Haryanto, E. V. (2011). Rancang Bangun Prototype Mesin Pencari String Menggunakan Metode Fuzzy String Matching. *Konferensi Nasional Sistem dan Informatika 2011*, 76-82.
- Hayati, R. (2020, Agustus 21). *Pengertian Skripsi, Ciri, Jenis, Tujuan, dan Cara Menulisnya*. Retrieved from penelitianilmiah.com: <https://penelitianilmiah.com/pengertian-skripsi/>
- Isodarus, P. B. (2017). Pembelajaran Bahasa Indonesia Berbasis Teks. *Jurnal Ilmiah Kebudayaan SINTESIS*, 1-11.
- Luthfi, K. &. (2009). *Algoritma Data Mining*. Yogyakarta: C.V ANDI OFFSET.
- Melita, R., Amrizal, V., Suseno, H. B., & Dirjam, T. (2018). PENERAPAN METODE TERM FREQUENCY INVERSE DOCUMENT FREQUENCY(TF-IDF) DAN COSINE

- SIMILARITY PADA SISTEM TEMU KEMBALI INFORMASI UNTUK MENGETAHUI SYARAH HADITS BERBASIS WEB (STUDI KASUS: SYARAH UMDATIL AHKAM). *JURNAL TEKNIK INFORMATIKA*, 149.
- Mulaab. (2017). *DATA MINING Konseo dan Aplikasi*. Malang: Media Nusa Creative.
- Nursyamsi. (2016). Kesalahan Penulisan Kata Bahasa Indonesia Dalam Jurnal Ilmiah Sains dan Teknologi. *Multilingual*, 193-204.
- Perkasa, T. R., Widyantara, H., & Susanto, P. (2014). Rancang Bangun Pendeteksi Gerak Menggunakan Metode Image Subtraction Pada Single Board Computer (Sbc). *Journal of Control and Network Systems*, 90-97.
- Putra, I. M., & Kusumawardani, R. P. (2017). ANALISIS TOPIK INFORMASI PUBLIK MEDIA SOSIAL DI SURABAYA MENGGUNAKAN PEMODELAN LATENT DIRICHLET ALLOCATION (LDA). *JURNALA TEKNIK ITS*, A311-A316.
- Rahadiyan Duwi Nugroho, C. T. (2018). Analisis Kesalahan dalam Penulisan Karya Ilmiah Mahasiswa Jepang dalam Pembelajaran BIPA. *Jurnal Pendidikan Bahasa dan Sastra*, 193-209.
- Ridlo, I. A. (2017). *Panduan Pembuatan Flowchart*. Surabaya: Fakultas Kesehatan Masyarakat Departemen Administrasi dan Kebijakan Kesehatan.
- Setijohatmo, U. T., Rachmat, S., Susilawati, T., & Rahman, Y. (2020). *Analisis Metoda Latent Dirichlet Allocation untuk Klasifikasi Dokumen Laporan Tugas Akhir Berdasarkan Pemodelan Topik*. Bandung: Prosiding The 11th Industrial Research Workshop and National Seminar.
- Syaifuddin, A., Alexandro, R., & Santoso, J. (2020). *Analisis Trending Topik untuk Percakapan Media Sosial dengan Menggunakan Topic Modelling Berbasis Algoritme LDA*. *JOURNAL OF INTELLIGENT SYSTEMS AND COMPUTATION*.
- Tanjung, B. S. (2018). Pendekatan Text Mining Sebagai Sistem Pendeteksi Pemberitahuan Palsu yang Tersebar dalam Twitter. *Seminar Nasional Ilmu Terapan (SNITER)*, 2.
- Urip T. Setijohatma, S. R. (2020). Analisis Metoda Latent Dirichlet Allocation untuk Klasifikasi Dokumen Laporan Tugas Akhir Berdasarkan Pemodelan Topik. *Prosiding The 11th Industrial Research Workshop and National Seminar Bandung, 26-27 Agustus 2020*, 1-7.
- Zulhanif. (2016). Pemodelan Topik dengan Latent Dirichlet Allocation. *Seminar Nasional Pendidikan Matematika 2016*, 1-8.
- Zulhanif, S. B. (2017). Aplikasi Latent Dirichlet Allocation (LDA) Pada Clustering Data Teks. *Jurnal Logika*, 1-6.