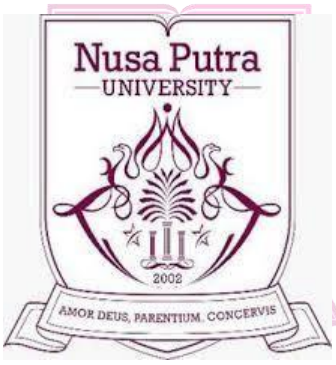
IMPLEMENTASI METODE *LATENT DIRICLHET ALLOCATION* (LDA) DALAM *TOPIC MODELING* PADA DATA BERITA KESEHATAN

(Studi Kasus Media *Online* Penyajian Berita Terpercaya Kompas.com)

SKRIPSI

YUNI YULISTIANTI

20200040146



PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK KOMPUTER DAN DESAIN

UNIVERSITAS NUSAPUTRA

SUKABUMI

JUNI 2024

IMPLEMENTASI METODE *LATENT DIRICLHET ALLOCATION* (LDA) DALAM *TOPIC MODELING* PADA DATA BERITA KESEHATAN

(Studi Kasus Media *Online* Penyajian Berita Terpercaya Kompas.com)

SKRIPSI

Diajukan Untuk Memenuhi Salah Satu Syarat

Dalam Menempuh Gelar Sarjana

Komputer Teknik Informatika

YUNI YULISTIANTI

20200040146

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK KOMPUTER DAN DESAIN

UNIVERSITAS NUSAPUTRA

SUKABUMI

JUNI 2024

PERNYATAAN PENULIS

JUDUL : Implementasi metode *Latent Dirichlet Allocation* (LDA) dalam *topic modeling* Pada Data berita Kesehatan (Studi Kasus Media *online* penyajian berita terpercaya Kompas.com)

NAMA : YUNI YULISTIANTI

NIM : 20200040146

"Saya menyatakan dan bertanggung jawab dengan sebenarnya bahwa Skripsi ini adalah hasil karya saya sendiri kecuali cuplikan dan ringkasan yang masing- masing telah saya jelaskan sumbernya. Jika pada waktu selanjutnya ada pihak lainyang mengklaim bahwa Skripsi ini sebagai karyanya, yang disertai dengan bukti- bukti yang cukup, maka saya bersedia untuk dibatalkan gelar Sarjana Komputer saya beserta segala hak dan kewajiban yang melekat pada gelar tersebut".

Sukabumi, 21 Juni 2024

YUNI YULISTIANTI

Penulis

PENGESAHAN SKRIPSI

JUDUL : IMPLEMENTASI METODE *LATENT DIRICHLET ALLOCATION* (LDA) DALAM *TOPIC MODELING* PADA DATA BERITA KESEHATAN (STUDI KASUS MEDIA *ONLINE* PENYAJIAN BERITA TERPERCAYA KOMPAS.COM

NAMA : YUNI YULISTIANTI

NIM : 20200040146

Skripsi ini telah diujikan dan dipertahankan di depan Dewan Penguji pada Sidang Skripsi tanggal 21 juni 2024 Menurut pandangan kami, Skripsi ini memadai dari segi kualitas untuk tujuan penganugerahan gelar Sarjana Komputer (S.Kom).

Sukabumi, 21 Juni 2024

Pembimbing I	Pembimbing II
Ivana Lucia Kharisma, M.Kom	**Alun Sujjada, S.Kom., M.T**
NIDN. 0429038002	NIDN. 0718108001

Ketua Penguji	Ketua Program Studi
Muhammad Ikhsan Thohir, M.Kom	**Ir. Somantri, S.T., M.Kom**
NIDN. 0415049302	NIDN. 0419128801

Plh Dekan Fakultas Teknik, Komputer dan Desain

Ir. Paikun S,T., M.T., IPM., ASEAN Eng
NIDN. 0402037401

PERUNTUKAN

Skripsi ini saya persembahkan kepada:

1. Tuhan Yang Maha Esa Allah SWT, atas rahmat dan hidayah-Nya yang selalu menyertai langkah hidup saya serta memberikan kekuatan dan kebijaksanaan selama proses penyusunan skripsi ini.
2. Ayah ibu dan juga adik tecinta yang selalu memberikan dukungan tanpa henti, baik secara moral maupun material. Terima kasih atas cinta, doa, dan pengorbanan yang tiada batas.
3. Dosen pembimbing yang telah memberikan bimbingan dan arahan selama penyusunan skripsi ini.
4. Pasangan tercinta yang selalu memberikan dukungan, cinta, dan semangat dalam setiap langkah perjalanan akademis ini. Terima kasih atas pengertian dan kesabaran yang luar biasa.
5. Teman-teman dan sahabat yang selalu memberikan semangat.

Terima kasih yang tulus dari hati terdalam untuk semua kontribusi dan dukungan yang diberikan. Semoga ini bermanfaat dan menjadi langkah awal untuk kontribusi yang lebih besar di masa depan.

ABSTRACT

The use of the internet and digital media has transformed news consumption from traditional sources to digital platforms like Facebook, Twitter, and Instagram, posing challenges related to the validity and reliability of information. News sites like kompas.com have adapted by offering multimedia formats and premium subscription services. In the context of health news, the complexity of topics and the variety of terms present challenges, necessitating detailed grouping. Latent Dirichlet Allocation (LDA) is an effective method for clustering health news and topic analysis. The research process involves problem identification, data collection, data preprocessing, LDA model creation, and evaluation. The LDA analysis of health news data identifies five main topics: Topic 0: body health (keywords: "signs", "alert", "attention", "body"). Topic 1: blood issues (keywords: "blood", "sugar"). Topic 2: disease symptoms (keywords: "know", "symptoms", "sick", "recover"). Topic 3: healthy drinks (keywords: "effects", "drink"). Topic 4: types of diseases (keywords: "blood", "sugar", "sleep", "eat", "fruit"). The implementation of LDA is conducted through a website using Streamlit, allowing for data upload, text preprocessing, and LDA application. A trial with new data addition and stemming removal shows that the analysis results are more accurate and relevant, retaining the original meaning of words and improving the quality of the analysis.

Keywords: Internet, Latent Dirichlet Allocation (LDA), topic modeling.

ABSTRAK

Penggunaan internet dan media digital telah mengubah konsumsi berita dari sumber tradisional ke platform digital seperti *Facebook, Twitter*, dan *Instagram*, menimbulkan tantangan terkait validitas dan reliabilitas informasi. Situs berita seperti kompas.com beradaptasi dengan format multimedia dan layanan berlangganan premium. Pada konteks berita kesehatan, kompleksitas topik dan variasi istilah menjadi tantangan, sehingga diperlukan pengelompokan yang terperinci. Metode *Latent Dirichlet Allocation (LDA)* efektif untuk pengelompokan berita kesehatan dan analisis topik. Tahapan penelitian melibatkan identifikasi masalah, pengumpulan data, pra-pemrosesan data, pembuatan model LDA, dan evaluasi. Hasil analisis LDA pada data kesehatan mengidentifikasi lima topik utama: Topik 0: kesehatan tubuh (kata kunci: "tanda", "waspada", "perhati", "tubuh"). Topik 1: masalah darah (kata kunci: "darah", "gula"). Topik 2: gejala penyakit (kata kunci: "kenal", "gejala", "sakit", "sembuh"). Topik 3: minuman sehat (kata kunci: "efek", "minum"). Topik 4: jenis penyakit (kata kunci: "darah", "gula", "tidur", "makan", "buah"). Implementasi LDA dilakukan melalui *website* menggunakan *streamlit*, memungkinkan pengunggahan data, pra-pemrosesan teks, dan penerapan LDA. Uji coba dengan penambahan data baru dan penghapusan *stemming* menunjukkan hasil analisis lebih akurat dan relevan, mempertahankan makna asli kata dan meningkatkan kualitas analisis.

Kata kunci : Internet, *Latent Dirichlet Allocation* (LDA), *topic modeling*.

KATA PENGANTAR

Puji syukur kami panjatkan ke hadirat Allah SWT, berkat Rahmat dan karunia-Nya akhirnya penulis dapat menyelesaikan skripsi yang berjudul "Implementasi metode *Latent Dirichlet Allocation* (LDA) dalam *topic modeling* Pada Data berita Kesehatan" Tujuan penulisan skripsi ini adalah untuk Mengidentifikasi topik-topik utama yang muncul dalam berita kesehatan, membantu pengguna menemukan berita dengan topik yang spesifik, serta Menyediakan wawasan yang lebih baik tentang tren dan perubahan topik yang sedang terjadi dalam berita kesehatan.

Sehubungan dengan itu penulis menyampaikan penghargaan dan ucapan terima kasih yang sebesar-besarnya kepada :

1. Dr. H. Kurniawan, ST., M.Si., MM., selaku Rektor Universitas Nusa Putra.
2. Bapak Anggy Pradiftha Junfithrana MT., selaku Wakil Rektor I Bidang Akademik Universitas Nusa
3. Bapak Somantri, S.T., M.Kom., selaku Ketua Program Studi Teknik Informatika Universitas Nusa Putra.
4. Ibu Ivana Lucia Kharisma, M.Kom. selaku pembimbing I yang telah meluangkan waktu dan tenaganya untuk membimbing dan memberikan arahan kepada penulis dalam menyelesaikan skripsi ini.
5. Bapak Alun Sujjada, S.Kom, M.T. selaku pembimbing II yang telah meluangkan waktu dan tenaganya untuk membimbing dan memberikan arahan kepada penulis dalam menyelesaikan skripsi ini.
6. Kepada diri sendiri Yuni Yulistianti : seorang penjelajah yang haus pengetahuan telah menembus berbagai tantangan rintangan dan ujian dalam hidup ini. Karya ini adalah bukti dari dedikasi dan semangat yang tak pernah padam, serta pengingat bahwa tidak ada yang mustahil bagi mereka yang berani bermimpi dan berjuang.

7. Sebagai kedua orang tua tercinta: Bapak Yusup dan Ibu Uun, dalam setiap detik perjuangan ini, kalian adalah sumber inspirasi dan kekuatan bagiku. Setiap tawa dan tangis, setiap pelukan hangat, telah menuntun langkahku menuju keberhasilan. Kalian adalah matahari dalam hari-hariku, menerangi setiap kegelapan dan memberi kehangatan dalam setiap dinginnya malam. Terima kasih atas setiap doa yang kalian panjatkan, atas setiap peluh yang kalian teteskan demi melihatku berhasil. Semua ini adalah bukti nyata dari cinta kalian yang abadi dan tulus. Tanpa kalian, aku takkan mampu berdiri sekuat ini, meraih mimpi dan menghadapi setiap tantangan dengan kepala tegak.
8. Sebagai adik tecinta Salwa Alawiyah: dalam setiap tawa dan tangismu, aku menemukan semangat untuk terus melangkah. Setiap kali kulihat senyum manismu, hatiku dipenuhi dengan harapan dan kebahagiaan. Kamu adalah cahaya kecil yang menerangi hari-hariku, memberikan warna dan makna dalam setiap langkah yang kuambil. Terima kasih adikku, atas segala dukungan dan pengertianmu. Dalam setiap perjuangan ini, kamu adalah penyemangat yang tak tergantikan. Walau kadang kita berbeda pendapat, tetapi kasih sayang di antara kita selalu lebih kuat dari segalanya. Kamu adalah anugerah terindah dalam hidupku, yang membuatku selalu berusaha menjadi contoh yang baik dan sosok yang bisa kamu banggakan.
9. Aji Amirudin terima kasih atas waktu, perhatian, dan bantuan yang selalu diberikan. Sosok luar biasa yang selalu hadir dalam berbagai situasi. Terima kasih atas waktu yang diluangkan serta saran, dukungan, dan bantuan yang tak pernah henti selama proses penyusunan skripsi ini. Bantuannya sangat berarti bagi penulis. Dukungan, cinta, dan semangat yang diberikan dalam setiap langkah perjalanan akademis ini sangat berharga. Terima kasih atas pengertian dan kesabaran yang luar biasa.
10. Anggia Putri Wulan Suci, Alyanissa Putri Iskandar, Putri Ayu Negara sebagai teman seperjuangan di masa perkuliahan. Terima kasih selalu berjuang bersama di keadaan sulit sekalipun.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Penulis mengharapkan kritik dan saran yang membangun demi perbaikan skripsi ini di masa mendatang. Semoga skripsi ini dapat bermanfaat bagi semua pihak yang membutuhkan. Akhir kata, semoga Allah SWT senantiasa memberikan rahmat dan hidayah-Nya kepada kita semua. Amin.

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIK

Sebagai aktivitas akademik UNIVERSITAS NUSA PUTRA, saya yang bertanda tangan dibawah ini :

Nama : Yuni Yulistianti

NIM : 20200040146

Program Studi : Teknik Informatika

Jenis Karya : Skripsi

Demi mengembangkan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Nusa Putra Hak Bebas Royalti Nonekslusif (Non-exclusive Royati-Free Right) atas karya saya yang berjudul :

IMPLEMENTASI METODE *LATENT DIRICLHET ALLOCATION* (LDA) DALAM *TOPIC MODELING* PADA DATA BERITA KESEHATAN (Studi Kasus Media *online* Penyajian Berita Terpercaya Kompas.com)

Berserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Nonekslusif di Universitas Nusa Putra berhak menyimpan, mengalih media/format-kan mengelola dalam bentuk pangkalan data (database), merawat dan mempublikasikan penelitian saya selama tetap mencantumkan nama saya sebagai penulis atas pencipta dan sebagai pemilik Hak Cipta.

Dibuat di : Sukabumi

Pada Tanggal : 21 Juni 2024

Yang menyatakan

Yuni Yulistianti

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR RUMUS

BAB I
PENDAHULUAN

1.1 Latar Belakang

Penggunaan Internet dan media digital mempunyai dampak yang signifikan terhadap industri berita. Hal ini terlihat dari perubahan pola konsumsi berita masyarakat yang beralih dari sumber tradisional ke *platform* digital. Internet telah mengubah cara orang mencari, mengonsumsi, dan berbagi berita. Perubahan besar terjadi dalam model bisnis industri berita khususnya media tradisional seperti surat kabar dan majalah yang mengalami penurunan, sementara itu *platform* digital semakin berkembang dan populer. Media sosial juga memengaruhi cara berita yang disebarkan [1]. Kebanyakan orang menerima berita melalui *platform* media sosial seperti Facebook, Twitter, dan Instagram. Hal tersebut mengubah dinamika sumber berita dan menimbulkan pertanyaan mengenai validitas dan reliabilitas informasi yang disebarluaskan [2]. Munculnya teknologi digital juga mengubah cara penyajian konten dalam industri berita, termasuk penggunaan multimedia, interaktivitas, dan personalisasi berita.

Salah satu situs berita terkemuka di Indonesia adalah kompas.com. Kompas.com menyajikan beragam berita mulai dari politik, ekonomi, olahraga, hiburan, dan lainnya, dengan jumlah artikel yang terus bertambah setiap harinya. Selain berita, kompas.com juga menawarkan beragam konten multimedia seperti artikel opini, laporan khusus, foto dan video yang memberikan pengalaman lebih mendalam bagi pembacanya. Kompas.com memiliki banyak rubrik khusus yang berfokus pada topik berita terkini, terpopuler, nasional, megapolitan, regional dan global. Situs tersebut juga menawarkan layanan berlangganan premium yang menyediakan konten eksklusif dan akses eksklusif ke berbagai berita terkini. Kompas.com aktif menyediakan liputan langsung peristiwa-peristiwa penting dan memiliki bagian komentar pembaca yang memungkinkan terjadinya interaksi antara pembaca dan penulis. Jumlah artikel yang terus bertambah setiap harinya, situs *website* ini menjadi sumber berita terpenting bagi masyarakat Indonesia dan dikunjungi oleh jutaan pembaca setiap bulannya.

Seiring bertambahnya jumlah artikel, pengguna semakin mencari cara yang efisien untuk menemukan informasi yang relevan dengan kebutuhan mereka, pengarsipan berita menjadi salah satu masalah yang krusial dengan melibatkan kurasi artikel dari berbagai topik seperti politik, ekonomi, olahraga, hiburan, kesehatan dan lainnya. Meskipun pengarsipan berita sudah cukup baik, namun masih diperlukan pengelompokan yang lebih terperinci untuk membantu pengguna menemukan berita dengan topik yang spesifik. Pada konteks pengelompokan berita kesehatan, terdapat beberapa permasalahan yang dihadapi, salah satunya adalah kompleksitas topik kesehatan yang luas dan bervariasi. Permasalahan lainnya yang dihadapi adalah fluktuasi topik dan variasi istilah dalam berita kesehatan, dikarenakan berita kesehatan dapat mencakup berbagai topik mulai dari penyakit spesifik hingga perkembangan terbaru dalam penelitian medis. Istilah yang digunakan untuk menggambarkan topik-topik ini juga dapat bervariasi secara luas. Hal ini dapat menjadi tantangan pengguna untuk mencari berita dengan topik yang spesifik sesuai dengan yang diinginkan [3].

Penggunaan informasi kesehatan yang tepat sangat penting dalam era digital saat ini. Informasi yang akurat dan *up-to-date* dapat membantu individu membuat keputusan yang lebih baik mengenai kesehatan mereka, termasuk informasi kesehatan dan fakta yang menunjukkan apakah benar bahwa masyarakat saat ini dapat dengan mudah menggunakan media informasi kesehatan tersebut hal ini tentu tidak terlepas dari beragamnya media informasi kesehatan yang digunakan oleh setiap kalangan masyarakat. Akses ke informasi kesehatan yang andal juga dapat meningkatkan kesadaran masyarakat tentang isu-isu kesehatan, memungkinkan mereka untuk mengambil langkah-langkah proaktif dalam menjaga kesejahteraan mereka. Meningkatnya jumlah informasi yang tersedia, tantangan utama adalah menyaring dan menemukan informasi yang relevan dan terpercaya, oleh karena itu pengelolaan dan pemanfaatan informasi kesehatan secara efektif menjadi semakin krusial untuk memastikan bahwa masyarakat dapat mengakses dan menggunakan informasi tersebut dengan benar [4].

Pada konteks pengelompokan berita kesehatan, lima topik yang akan ditambahkan untuk membantu pengguna menemukan informasi yang relevan adalah: kesehatan tubuh, masalah darah dan ciri penyakit, gejala penyakit, efek minuman, dan juga jenis penyakit. Mengambil lima topik pengelompokan berita kesehatan yang telah disebutkan sebelumnya, pengguna akan dapat dengan lebih mudah menavigasi informasi kesehatan yang kompleks dan bervariasi. Masalah kompleksitas topik kesehatan yang luas dan variasi istilah yang digunakan dalam berita kesehatan semakin menegaskan perlunya pengelompokan yang terperinci untuk membantu mengatasi tantangan pengguna dalam menavigasi informasi kesehatan yang beragam dan seringkali kompleks serta dengan adanya fokus pada topik-topik tersebut, diharapkan pengguna dapat dengan lebih efisien menemukan berita yang sesuai dengan kebutuhan dan minat kesehatan mereka [5].

Kompas.com dipilih sebagai sumber data dalam penelitian ini karena merupakan salah satu situs berita terkemuka di Indonesia dengan reputasi yang baik dan jangkauan pembaca yang luas. Selain itu, Kompas.com memiliki beragam kategori berita yang mencakup banyak topik, termasuk kesehatan, sehingga menyediakan basis data yang kaya untuk analisis. Kompas.com juga memiliki karakterisktik ruang yang tidak terbatas, khalayak dapat memilih beritanya sendiri, berita berdiri sendiri sehingga khalayak tidak harus membaca berita secara berurutan, berita di Kompas.com tersimpan dan bisa diakses kembali kapan pun, berita disampaikan dengan sangat cepat dan langsung, kemampuan multimedia, dan interaktivitas antara redaksi dengan pembaca. Data kesehatan dari Kompas.com dipilih karena situs ini menyajikan berita kesehatan yang komprehensif dan terkini, mencakup berbagai aspek kesehatan. Dengan menggunakan data dari sumber yang tepercaya seperti Kompas.com, diharapkan hasil penelitian ini akan lebih relevan dan bermanfaat bagi pengguna yang mencari informasi kesehatan [6].

Salah satu metode yang digunakan untuk pemodelan topik adalah *Latent Dirichlet Allocation* (LDA). LDA adalah sebuah model generatif yang

digunakan dalam pemodelan topik dalam teks, LDA digunakan untuk menemukan topik yang tersembunyi dalam koleksi dokumen. Ada beberapa riset terkait seperti penelitian Blei, Ng, & Jordan memperkenalkan metode *Latent Dirichlet Allocation* (LDA) untuk melakukan analisis topik pada koleksi dokumen. Metode ini telah banyak digunakan dalam pengelompokan berita dan analisis topik dalam berbagai domain [7]. Penelitian Fitri Bimantoro sebuah model yang menggabungkan metode LDA dengan jaringan saraf tiruan (*neural network*) mengenai pengenalan pola aksara jawa mengenali pola tulisan tangan aksara sasak sehingga kedepannya kedua metode ini dapat dikembangkan untuk sistem pembelajaran aksara sasak [8]. Nuraisa Novia Hidayati dkk yang membandingkan metode LDA dan NMF untuk mengetahui metode mana yang memiliki hasil terbaik. Hasilnya menunjukkan bahwa LDA memiliki kinerja yang lebih baik daripada NMF sementara itu, pembentukan data pelatihan yang lebih baik adalah diperlukan untuk membuat model NER yang tidak *overfitting* [9]. Riset-riset ini menunjukkan bahwa metode *machine learning* khususnya metode LDA, memiliki potensi besar dalam pengelompokan berita kesehatan dan analisis topik dalam konteks berita kesehatan.

Berdasarkan latar belakang permasalahan tersebut maka penulis mempunyai gagasan penelitian yaitu dengan mengangkat **"IMPLEMENTASI METODE *LATENT DIRICLHET ALLOCATION* (LDA) DALAM *TOPIC MODELING* PADA DATA BERITA KESEHATAN".**

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas maka perumusan masalah dalam penelitian ini yaitu :

1. Bagaimana mengimplementasikan metode *latent dirichlet allocation* (LDA) pada data berita kesehatan kompas.com?
2. Bagaimana metode *latent dirichlet allocation* (LDA) dapat digunakan untuk mengidentifikasi topik berita kesehatan kompas.com?
3. Bagaimana implementasi *website topic modeling* menggunakan *streamlit*

dapat memudahkan pengguna dalam menemukan beragam topik pada berita kesehatan kompas.com?

1.3 Batasan Masalah

Berdasarkan latar belakang masalah di atas, penulis menggunakan batasan penulisan agar di dalam pembahasan dan isi yang ada di dalam penulisan ini tidak melebar dan menyimpang dari judul. Adapun batasan – batasan yang diberikan adalah sebagai berikut :

1. Data berita kesehatan dibatasi hanya pada data yang dapat diakses dan dikumpulkan, atau data yang sudah tersedia didalam Kompas.com dari tahun Januari 2023 - Februari 2024 sekarang.
2. Data yang digunakan mencakup judul berita, isi berita, dan tanggal publikasi.
3. Artikel berita yang diambil dalam bahasa Indonesia.
4. Berita kesehatan akan dikelompokkan dalam 5 kategori yaitu: kesehatan tubuh, masalah darah dan ciri penyakit, gejala penyakit, efek minuman, dan juga jenis penyakit.

1.4 Tujuan dan Manfaat Penelitian

Tujuan dari Implementasi metode *Latent Dirichlet Allocation* (LDA) dalam topik modeling Pada berita kesehatan dengan topik penelitian Kompas.com adalah untuk membantu dalam analisis dan pengelompokan berita kesehatan dengan topik yang relevan. Berikut adalah tujuan dan manfaat penelitian ini:

Tujuan:

1. Mengidentifikasi metode *latent dirichlet allocation* (LDA).
2. Mengimplementasikan metode *latent dirichlet allocation* (LDA) pada data berita kesehatan kompas.com.
3. Mengimplementasikan *website topic modeling* menggunakan *streamlit* untuk membangun sistem yang menyajikan beragam topik pada berita kesehatan kompas.com.
4. Mengembangkan metode pengelompokan berita kesehatan yang otomatis dan terukur, serta memungkinkan untuk mengeksplorasi

konten dengan lebih efisien.

Manfaat:

Manfaat yang akan didapatkan dari penelitian ini adalah:

1. Membantu untuk memperoleh pemahaman yang lebih mendalam terhadap beragam topik yang muncul dalam berita kesehatan.
2. Membantu menemukan berita kesehatan dengan beragam topik yang spesifik melalui *website topic modeling* menggunakan *streamlit.*
3. Pengelompokkan kategori data berita dapat dilakukan secara otomatis, sehingga waktu yang digunakan semakin efisien.

1.5 Sistematika Penulisan

Memberikan gambaran secara garis besar, dalam hal ini dijelaskan isi dari masing masing bab dari penelitian ini. Sistematika penulisan dalam pembuatan laporan ini sebagai berikut :

BAB I : PENDAHULUAN

Dalam bab ini dibahas mengenai: Latar Belakang Masalah, Batasan Masalah, Rumusan Masalah, Tujuan dan Manfaat Penelitian, dan Sistematika Penelitian.

BAB II : TINJAUAN PUSTAKA

Dalam bab ini dibahas mengenai: Penelitian terkait, Landasan Teori dan Kerangka Pemikiran.

BAB III : METODOLOGI PENELITIAN

Dalam bab ini dibahas mengenai:Tahapan penelitian yang dilakukan, pengumpulan data serta perancangan sistem.

BAB IV Hasil dan Pembahasan

Merupakan bagian penting dalam sebuah penelitian. Hasil adalah jawaban dari pertanyaan penelitian yang dituliskan di bagian pendahuluan. Sedangkan pembahasan adalah bagian di mana hasil tersebut didiskusikan dan dianalisis lebih lanjut, serta dibandingkan

dengan teori atau penemuan sebelumnya.

BAB V Saran dan Kesimpulan

Didasarkan pada hasil analisis dan interpretasi data yang telah dikumpulkan. Saran-saran diberikan sebagai pertimbangan untuk membantu penyelenggaraan penelitian dengan lebih baik.

B V
PENUTUP

B
A

Topic Modeling menggunakan *Latent Dirichlet Allocation* (LDA) adalah teknik yang digunakan untuk mengidentifikasi topik utama yang muncul dalam kumpulan dokumen teks. Tujuannya adalah untuk menemukan pola tersembunyi dalam teks yang memungkinkan dokumen tersebut dikelompokkan ke dalam topik-topik yang berbeda. Langkah pertama adalah menginstal alat yang diperlukan, seperti *sklearn*, Sastrawi, dan perpustakaan *Natural Language Processing (NLP) Python*. Alat-alat ini membantu dalam mengubah data ke dalam format terstruktur dan meningkatkan kinerja model data. Langkah kedua melibatkan *import library* seperti *pandas, numpy, NLTK, RE (Regular Expression),* sastrawi, dan kerangka data *pandas python*. *Matplotlib.pyplot* digunakan untuk visualisasi data, dan *Seaborn* digunakan untuk visualisasi data yang lebih efisien. Pra-pemrosesan data untuk LDA melibatkan beberapa langkah, termasuk *case folding*, *tokenizing, stopwords*, *stemming*, dan pemodelan topik.

Hasil dari pemodelan topik LDA pada berita kesehatan diimplementasikan melalui website menggunakan Streamlit, yang memungkinkan pengguna untuk mengunggah data berita kesehatan, melakukan *preprocessing* pada teks, mengekstrak fitur menggunakan metode TF-IDF, dan menerapkan LDA untuk mendapatkan topik-topik dari teks. Visualisasi model LDA menggunakan *Word Count* dan *Wordcloud* juga digunakan. Evaluasi model LDA menggunakan *elbow* dan *silhotte score* penting untuk memastikan model topik memberikan informasi yang akurat dan komprehensif tentang struktur topik dalam data kesehatan dari kompas.com. Melalui antarmuka website, pengguna dapat mengikuti langkah-langkah yang disediakan untuk mendapatkan hasil analisis topik dari data tersebut.

Hasil dari analisis topik menggunakan metode LDA pada data kesehatan mengidentifikasi lima topik, masing-masing dengan kata kunci yang mewakili topik tersebut beserta bobotnya. Topik 0 berkaitan dengan berbagai aspek kesehatan tubuh seperti berat badan, makanan pantang, dampak obat, dan upaya

menjaga kesehatan tubuh secara umum, dengan kata kunci utama "tanda", "waspada", "perhati", dan "tubuh". Topik 1 fokus pada masalah darah seperti kadar gula darah, ciri-ciri penyakit tertentu, hubungan dengan kehamilan, serta informasi tentang seks dan anak-anak, dengan kata kunci utama "darah" dan "gula". Topik 2 menyoroti gejala-gejala penyakit dan tanda-tanda kesehatan lainnya, termasuk jenis-jenis penyakit, perhatian terhadap tubuh, olahraga, dan manfaat vitamin, dengan kata kunci utama "kenal", "gejala", "sakit", dan "sembuh". Topik 3 berkaitan dengan minuman seperti kopi dan teh serta efeknya terhadap kesehatan, baik manfaat maupun efek sampingnya, dengan kata kunci utama "efek" dan "minum". Topik 4 fokus pada berbagai jenis penyakit seperti jantung, diabetes, ginjal, dan kanker, serta bahaya kesehatan seperti stroke dan risiko hipertensi, dengan kata kunci utama "darah", "gula", "tidur", "makan", dan "buah". Analisis topik ini memberikan gambaran tentang berbagai aspek kesehatan yang dibahas dalam dataset, mulai dari masalah berat badan dan makanan hingga masalah darah, gejala penyakit, minuman sehat, dan berbagai jenis penyakit yang perlu diwaspadai. Uji coba dengan menamabahkan data baru dan penghapusan *stemming* juga dilakukan untuk memperoleh hasil yang diperoleh menjadi lebih akurat dan relevan serta untuk menentukan dampaknya terhadap kualitas dan ketepatan hasil analisis keseluruhan hasilnya adalah hasil analisis setelah uji coba lebih baik dalam hal kualitas dan ketepatan. Kata-kata kunci yang muncul lebih relevan dengan topik kesehatan yang sedang dianalisis, dan penghapusan *stemming* membantu mempertahankan makna asli kata-kata tersebut.

5.1 Saran

Pada penelitian ini, beberapa temuan penting telah diidentifikasi yang memberikan kontribusi penting terhadap pemahaman penulis tentang *latent dirichlet allocation*. Temuan ini menyoroti pentingnya pemodelan topik dalam konteks pengelompokkan berita kesehatan dan menawarkan wawasan yang berharga bagi praktisi, akademisi, dan pembuat kebijakan. Pada penelitian ini juga memiliki keterbatasan, keterbatasan utamanya adalah dari evaluasi dan saat mengimplementasikannya hasil kedalam *website* menggunakan *freamwork streamlit* ini adalah kurangnya pemahaman yang mendalam dari penulis terhadap

metode evaluasi yang digunakan dan terkait *streamlit* tersebut. Hal ini dapat mengarah pada penilaian yang tidak konsisten dan mungkin tidak sepenuhnya mencerminkan kualitas sebenarnya dari model topik yang dihasilkan.

Sebagai saran untuk penelitian selanjutnya, disarankan agar peneliti lebih memahami terkait *streamlit* dan metode evaluasi, sehingga hasilnya menjadi lebih konsisten, akurat, dan dapat dipercaya. Terakhir, penulis ingin menyampaikan terima kasih kepada semua pihak yang telah mendukung penelitian ini, terutama kepada pembimbing, ayah serta ibu, dan juga teman-teman. Tanpa dukungan mereka, penelitian ini tidak mungkin tercapai. Semoga penelitian ini dapat memberikan kontribusi yang berharga bagi pengembangan ilmu pengetahuan dan praktik di masa depan. Sekali lagi, terima kasih atas kesempatan ini dan harapan saya agar penelitian ini dapat memberikan manfaat yang signifikan bagi pembaca.

DAFTAR PUSTAKA

[1] Lavanya Rajendran and Preethi Thesinghraja “The Impact of New Media on Traditional Media” 2014.

[2] Nic Newman with Richard Fletcher, Anne Schulz, Simge Andı, Craig T. Robertson, and Rasmus Kleis Nielsen “Reuters Institute Digital News Report”, 2021.

[3] repository.uin-suska.ac.id/15486/9/9.%20BAB%20IV_2018151KOM

[4] Ditha Prasanti, "Potret Media Informasi Kesehatan Bagi Masyarakat Urban di Era Digital", IPTEK-KOM, Vol. 19 No. 2, Desember 2017: 149-162.

[5] Nurina Savanti Widya Gotami, Indriati, Ratih Kartika Dewi “Peringkasan Teks Otomatis Secara Ekstraktif Pada Artikel Berita Kesehatan Berbahasa Indonesia Dengan Menggunakan Metode Latent Semantic Analysis” 2018.

[6] Halimatul Abkoriyah , Tribuana Tungga Dewi, “OBJEKTIVITAS BERITA DI HARIANKOMPAS DAN KOMPAS.COM (ANALISIS ISI PEMBERITAAN KASUS PEMBUNUHAN ENGELINE)”, Journal of Strategic CommunicationVol. 7, No. 2, Hal. 40-53 Maret 2017.

[7] David M. Blei, Andrew Y. Ng, Michael I. Jordan “Latent Dirichlet Allocation” Journal of Machine Learning Research 3 (2003) 993-1022.

[8] A.A.Sg.Mas Karunia Maharani, Fitri Bimantoro, “PENGENALAN POLA TULISAN TANGAN AKSARA SASAK MENGGUNAKAN METODE LINEAR DISCRIMINANT ANALYSIS DAN JARINGAN SYARAF TIRUAN JENIS BACKPROPAGATION” JTIKA, Vol. 2, No. 2, September 2020.

[9] Nuraisa Novia Hidayati, Putri Damayanti, Agus Zainal Arifin, Maryamah, Rarasmaya Indraswari, Rizka Wakhidatus Sholikah, “Identification of Traffic Information on Twitter Data using Topic Modeling and Named Entity Recognition”, JLK Vol 4, No 1 Maret 2021.

[10] M. LUVIAN CHISNI CHILMI, " LATENT DIRICHLET ALLOCATION (LDA) UNTUK MENGETAHUI TOPIK PEMBICARAAN WARGANET TWITTER TENTANG OMNIBUS LAW", UNIVERSITAS ISLAM NEGERI SYARIF HIDAYATULLAH, 2021.

[11] Wahyudin, "APLIKASI TOPIC MODELING PADA ANALISIS PEMBERITAAN OLEH PORTAL BERITA ONLINE SELAMA MASA PSBB PERTAMA".

[12] Agustina Lili, Suhada, Saputra Widodo, "Pengelompokan Hasil Panen Kelapa Sawit Dalam Produksi Per Blok Menggunakan Algoritma K-Means" Journal of Machine Learning and Data Analytics (MALDA) Volume 01, No.01, Januari 2022 Page: 45-54.

[13] Mayadi, Siti Setiawati, Wowon Priatna, "Pengelompokan Hasil Survei MBKM Menggunakan K-Mean dan K-Medoids Clustering" JURNAL MEDIA INFORMATIKA BUDIDARMA Volume 7, Nomor 1, Januari 2023, Page 426-435.

[14] Siti Sarah, Mustakim, "Analisis Penerimaan Vaksin Covid-19 Berbasis Fuzzy Clustering Machine Learning di Provinsi Riau" JURIKOM (Jurnal Riset Komputer), Vol. 8 No. 6, Desember 2021.

[15] Kamdan, Ivana Lucia Kharisma, Gina Purnama Insany, Paikun, "Research Topic Modeling in Informatics Engineering Study Program at Nusa Putra University using LDA method" International Journal of Engineering and Applied Technology (IJEAT) ISSN 2620-9632 Vol. 5., No. 2, November 2022, pp. 24-35.

[16] Rimbun Siringoringo, Jamaluddin, Resianta Perangin-Angin, "PEMODELAN TOPIK BERITA MENGGUNAKAN LATENT DIRICHLET ALLOCATION DAN K-MEANS CLUSTERING" Jurnal Informatika Kaputama(JIK), Vol. 4 No. 2 Juli 2020.

[17] Bagas Dwi Santosa , Nurul Fatimah , Netania Indi Kusumaningtyas , Ulfi Saidata Aesyi , Herdiesel Santoso, "Analisis Kepercayaan Masyarakat Tentang Kepolisian Indonesia di Twitter Menggunakan Latent Dirichlet Allocation (LDA)"

Indonesian Journal on Data Science ISSN 2987-7423 - Vol. 1, No. 2, November 2023, hlm 66-76.

[18] Cheng-Hsuan Li, Bor-Chen Kuo, Member, IEEE, and Chin-Teng Lin, Fellow, IEEE "LDA-Based Clustering Algorithm and Its Application to an Unsupervised Feature Extraction" 2011.

[19] Satria Ardi Perdana, Sara Famayla Florentin , dan Agus Santoso "ANALISIS SEGMENTASI PELANGGAN MENGGUNAKAN K-MEANS CLUSTERING STUDI KASUS APLIKASI ALFAGIFT" 2022.

[20] Muhammad Rheza Palevi, Zulfahmi Indra, "Implementasi Algoritma K-Means Clustering Dengan Pendekatan Active Learning Pada Siswa SMA Untuk Menentukan Jurusan Ke Perguruan Tinggi" Jurnal SAINTIKOM (Jurnal Sains Manajemen Informatika dan Komputer) Volume 23 ; Nomor 1 ; Februari 2024 ; Page 26-36.

[21] Muhammad Rijal Fadli, " Memahami desain metode penelitian kualitatif " , Humanika, Kajian Ilmiah Mata Kuliah Umum, ISSN: 1412-1271 (p); 2579-4248 (e). Vol. 21. No. 1. (2021). pp. 33-54 doi: 10.21831/hum.v21i1. 38075. 33-54.

[22] Alif Iffan Alfanzar, Khalid, Indri Sudanawati Rozas "TOPIC MODELLING SKRIPSI MENGGUNAKAN METODE LATENT DIRICLHET ALLOCATION" 2020.

[23] Milad Rogha , Subham Sah, Alireza Karduni , Douglas Markant, and Wenwen Dou, "The Impact of Elicitation and Contrasting Narratives on Engagement, Recall and Attitude Change with News Articles Containing Data Visualization" 2024.

[24] Si Chen and Yufei Wang Department of Electrical and Computer Engineering University of California San Diego "Latent Dirichlet Allocation"

[25] Faza Rashif1, Goldio Ihza Perwira Nirvana2, Muhammad Alif Noor3, Nur Aini Rakhmawati "Implementasi LDA untuk Pengelompokan Topik Cuitan Akun Bot Twitter bertagar #Covid-19" 2021.

[26] Chairullah Naury, “Topic Modelling pada Sentimen terhadap Headline Berita Online Berbahasa Indonesia” Universitas Islam Indonesia, 2020.

[27] Hamed Jelodar, Yongli Wang, Chi Yuan, Xia Feng, Xiahui Jiang, Yanchao Li Liang Zhao, “Latent Dirichlet Allocation (LDA) and Topic modeling: models, applications, a survey” 2018.

[28] Putu Manik Prihatini, I Ketut Suryawan, I Nyoman Mandia, “METODE LATENT DIRICHLET ALLOCATION UNTUK EKSTRAKSI TOPIK DOKUMEN” JURNAL LOGIC. VOL. 17. NO. 3. NOPEMBER 2017.

[29] Dziky Ridhwanullah, Dhomas Hatta Fudholi, “Pemodelan Topik pada Cuitan tentang Penyakit Tropis di Indonesia dengan Metode Latent Dirichlet Allocation”, Jurnal Ilmiah Sinus (JIS) Vol : 20, No. 1, Januari 2022.

[30] Hery Oktafiandi, “Implementasi LDA untuk Pengelompokan Topik Twitter Bertagar #Mypertamina”, 2023.