

ANALISIS DAN PENERAPAN TOPIC MODELING PADA JUDUL TUGAS AKHIR MAHASISWA MENGGUNAKAN METODE LATENT DIRICHLET ALLOCATION (LDA)

Titiana*¹⁾, Dwi Hosanna Bangkalang²⁾

1. Universitas Kristen Satya Wacana, Salatiga, Indonesia
2. Universitas Kristen Satya Wacana, Salatiga, Indonesia

Article Info

Kata Kunci: Tugas Akhir; LDA; *Topic Modeling*; *Cluster*; *Hyperparameter Tuning*

Keywords: *Final Project*, LDA, *Topic Modeling*, *Cluster*, *Hyperparameter Tuning*

Article history:

Received 18 May 2023

Revised 1 June 2023

Accepted 15 June 2023

Available online 1 December 2023

DOI :

<https://doi.org/10.29100/jipi.v8i4.4254>

* Corresponding author.

Titiana

E-mail address:

titiana262001@gmail.com

ABSTRAK

Tugas akhir menjadi syarat kelulusan bagi mahasiswa program studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana. Penentuan judul tugas akhir di program studi, biasanya ditentukan sesuai konsentrasi. Namun, saat penentuan judul masih terdapat mahasiswa yang kebingungan karena tidak mempunyai referensi yang valid terkait gambaran judul di konsentrasi program studi. Sehingga, mahasiswa sering memasukkan data kosong pada judul tugas akhir saat melakukan pendaftaran. Selain itu, program studi juga belum mempunyai informasi terkait pengelompokan judul tugas akhir mahasiswa. Oleh karena itu dibutuhkan *topic modeling* untuk mengetahui minat judul tugas akhir sebagai referensi mahasiswa dalam penentuan judul. *Topic modeling* dibuat menggunakan metode Latent Dirichlet Allocation (LDA) kemudian dibuat pengklasteran serta dilakukan evaluasi model menggunakan *hyperparameter tuning* dan *coherence score* sebagai alat ukur performa model. Tahapan penelitian meliputi studi literatur dan wawancara, pengumpulan dan analisis data, *data pre-processing*, pembuatan model, dan evaluasi. Diperoleh hasil bahwa mayoritas mahasiswa Sistem Informasi mengambil judul bertema Tata Kelola. Model yang dibuat memperoleh 4 *cluster* yaitu terkait *Technology Acceptance Model*, *Framework Cobit*, dan *Arsitektur Enterprise*. Setelah dilakukan *hyperparameter tuning* didapat *coherence score* sebesar 0,617789 dari yang sebelumnya 0,53448.

ABSTRACT

The final project is a requirement for graduation for students in the Information Systems program, Faculty of Information Technology, Satya Wacana Christian University. The determination of the final project title in the program is usually based on the concentration. However, during the title determination process, some students are confused because they do not have valid references regarding the description of titles in their program's concentration. In addition, the program also lacks information regarding the grouping of students' final project titles. Therefore, topic modeling is needed to determine the interest in final project titles as a reference for students in determining their titles. Topic modeling is created using the Latent Dirichlet Allocation (LDA) method, followed by clustering and model evaluation using hyperparameter tuning and coherence score as performance metrics. The research stages include literature review and interviews, data collection and analysis, data pre-processing, model creation, and evaluation. The results obtained four clusters related to the Technology Acceptance Model, Cobit Framework, and Enterprise Architecture with the theme of Governance. The model created achieved a coherence score of 0.617789 after hyperparameter tuning, compared to the previous score of 0.53448.

I. PENDAHULUAN

PROGRAM Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana setiap tahun meluluskan mahasiswa yang sudah menyelesaikan tugas akhir. Tugas akhir adalah salah satu syarat yang harus diselesaikan oleh mahasiswa dalam menyelesaikan pendidikan di perguruan tinggi tertentu yang

merupakan kajian dari suatu keadaan yang diungkap secara ilmiah[1]. Esensi dari tugas akhir sesungguhnya adalah salah satu cara mengetahui pengetahuan dari seorang mahasiswa secara individual setelah selesai menempuh seluruh matakuliah, agar ketika menyandang gelar sarjana hal tersebut dapat dipertanggungjawabkan kepada institusi dan masyarakat. Sehingga dalam upaya menjaga kualitas mahasiswa yang akan diluluskan dan akan bergabung ke tengah-tengah masyarakat, maka dari itu setiap mahasiswa wajib membuat tugas akhir[2].

Program studi Sistem Informasi memiliki dua profil lulusan yaitu Bisnis Analis dan Audit Tata Kelola Teknologi Informasi yang diadopsi dari Association For Information Systems Indonesia (AISINDO). Sehingga konsentrasi menjadi bidang fokus nya dan profil lulusan menjadi *role* nya. Dengan jumlah mahasiswa yang banyak, Program studi Sistem Informasi terbagi menjadi dua konsentrasi yaitu, Tata Kelola dan Sistem Informasi Geografis. Konsentrasi merupakan fokus mahasiswa terhadap suatu bidang studi tertentu sesuai dengan minatnya agar mahasiswa difokuskan terhadap ilmu yang didapat dari matakuliah sebelumnya, sehingga dapat lebih terarah dan dapat menentukan judul tugas akhir sesuai minat mahasiswa[3].

Namun, terdapat beberapa kendala saat mahasiswa menentukan judul tugas akhir. Pertama, mahasiswa yang masih kesulitan dalam penentuan judul karena tidak memiliki referensi pengambilan judul tugas akhir, sehingga mengakibatkan salah satu faktor *timeline* pembuatan tugas akhir dapat terganggu. Kedua, ketidaktahuan akan pengambilan judul tugas akhir, maka mahasiswa akan memasukan data kosong ke dalam Sistem Informasi tugas akhir yang dimiliki oleh Fakultas yang mengakibatkan bagian Program studi menjadi kesulitan untuk mendistribusikan dosen pembimbing tugas akhir karena bagian Program studi tidak mendapatkan gambaran Judul tugas akhir yang akan dibuat oleh Mahasiswa, sedangkan dalam pendistribusian dosen pembimbing dibutuhkan gambaran Judul tugas akhir Mahasiswa agar dapat disesuaikan dengan kepakaran dosen. Ketiga, karena ketidaksesuaian kepakaran dosen saat pendistribusian dosen pembimbing, terdapat mahasiswa yang mengajukan pergantian dosen pembimbing sehingga mengakibatkan pengulangan proses administrasi yang tidak efisien.

Untuk mengatasi permasalahan tersebut, dibutuhkan *topic modeling* pada pengambilan Judul tugas akhir Mahasiswa Program studi Sistem Informasi. *topic modeling* merupakan suatu pendekatan untuk melakukan pengelompokan data yang berupa teks berdasarkan suatu topik tertentu dari sebuah dokumen dan pendekatan tersebut masuk dalam pendekatan *clustering* dalam studi *machine learning*[4]. *Topic modeling* dapat membantu dalam melakukan indentifikasi dan menghasilkan tema-tema penting dalam koleksi data tekstual Judul tugas akhir Mahasiswa Program studi Sistem Informasi yang tidak terstruktur dan sulit diidentifikasi oleh manusia secara manual untuk mendapatkan tema tersembunyi dan menemukan topik dari teks yang jumlahnya besar[5][6][7].

Metode yang digunakan untuk melakukan pendekatan *topic modeling* sangat beragam, penelitian yang dilakukan oleh Mailoa, Damayanti, dan Christopher[8] yang membahas mengenai *topic modeling* dan *text clustering* menggunakan metode *hierarchical agglomerative clustering* pada data sebuah Program studi berhasil mengetahui jumlah dan judul tugas akhir yang muncul di tiap *cluster*, memperoleh judul yang sering muncul atau digunakan oleh mahasiswa. Penelitian yang dilakukan oleh Alif, Khalid, dan Indri[9] yang membuat pemodelan topik skripsi pada suatu program studi menggunakan teknik Latent Dirichlet Allocation (LDA) karena LDA dianggap mampu meringkas, mengklusterkan, menghubungkan, dan memproses data teks yang sangat besar. Dengan dilakukan banyak percobaan, hasil *cluster* tersebut telah diverifikasi oleh pihak stakeholder Program studi Sastra Inggris (UINSA) bahwa kata-kata yang ada pada topik *cluster* sesuai dengan menurut konsentrasi pada Program studi Sastra Inggris (UINSA).

Pada penelitian yang dilakukan oleh Nathanael, Siti, dan Amalia[10] menggunakan Latent Dirichlet Allocation untuk pemodelan topik tugas akhir mahasiswa Fakultas Institut Teknologi Telkom Purwokerto tahun 2022. Menghasilkan 8 topik dengan nilai *coherence* sebesar 0.446752 setelah dilakukan 20 kali percobaan, teks yang berhasil dihasilkan setelah dilakukan pemodelan serta evaluasi sudah menggambarkan topik yang diambil. Pada penelitian yang dilakukan oleh Aksin dan Moch Ari[3] membahas *topic modeling* pada penelitian dosen di Google Scholar menggunakan pendekatan Latent Dirichlet Allocation untuk mencari kecenderungan judul penelitian yang dilakukan oleh dosen JPTEI UNY digunakan *coherence value* sebagai alat ukur dalam menentukan jumlah topik dan didapat hasil *coherence value* sebesar 0,4076 dan peneliti mengambil 4 topik sebagai hasil yang optimum untuk memodelkan judul penelitian dengan baik.

Pada penelitian yang dilakukan oleh Kherwa dan Bansal[11] yang membahas *topic modeling review* melakukan perbandingan teknik Latent Semantic Analysis(LSA) dan Latent Dirichlet Allocation (LDA) mendapatkan hasil yaitu nilai *coherence* yang dihasilkan LDA lebih tinggi dari LSA. Maka dari itu, LDA dianggap cukup mampu untuk menjadi alat yang efektif untuk analisis statistik koleksi dokumen. Meskipun LDA efektif digunakan dalam klasifikasi, pemilihan fitur, dan pencarian informasi, namun menggunakan LDA pada kumpulan data besar

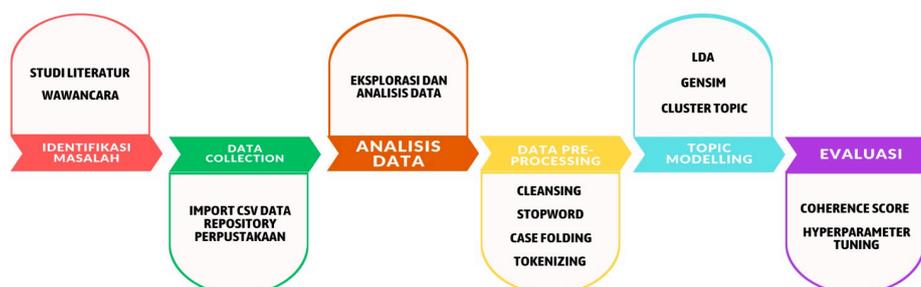
membutuhkan waktu berjam-jam bahkan berhari-hari karena komputasi yang berat dan akses memori yang intensif.

Pada penelitian yang dilakukan oleh Naury, Fudholi, dan Hidayatullah[12] yang membahas *topic modeling* pada sentimen *headline* berita *online* menggunakan metode Latent Dirichlet Allocation (LDA) dan metode Long Short-term Memory (LSTM) dapat memberikan topik-topik yang telah dibicarakan di media *online*, untuk topik yang memiliki nilai sentimen positif, negatif, dan netral. Topik - topik yang berhasil diperoleh dari pemodelan pada berita *online* yang digunakan sebagai objek penelitian ini dapat diinterpretasikan dengan baik.

Dari banyaknya rujukan peneliti terdahulu dalam menggunakan metode Latent Dirichlet Allocation (LDA), maka dari itu penelitian ini akan menggunakan metode LDA dalam pembuatan *topic modelling* dengan perbedaan di bagian cara mengevaluasi model dari penelitian serupa yang membahas pembuatan *topic modeling* pada judul tugas akhir mahasiswa yaitu proses *hyperparameter tuning* yang akan dijelaskan di bagian metode penelitian dan pembahasan. Sehingga dapat dirumuskan masalah yang pertama, bagaimana menerapkan *topic modeling* menggunakan metode LDA pada judul tugas akhir Mahasiswa. Kedua, bagaimana melakukan evaluasi terhadap model yang sudah dibuat dengan proses *hyperparameter tuning*. Tujuan penelitian ini adalah memodelkan topik tugas akhir mahasiswa sehingga dapat dijadikan informasi *cluster* pengambilan judul tugas akhir dari peneliti terdahulu atau mahasiswa lainnya. *Topic modeling* memungkinkan untuk menemukan informasi pengambilan judul tugas akhir Program studi Sistem Informasi[5]. Sehingga dapat bermanfaat sebagai referensi untuk mahasiswa yang akan membuat tugas akhir dan sehingga mahasiswa dapat mengetahui perencanaan judulnya dan memudahkan bagian Program studi dalam strategi pembagian dosen pembimbing kedepannya.

II. METODE PENELITIAN

Penelitian ini akan dilakukan urut mulai dari mengidentifikasi masalah, kemudian pengumpulan dan pemrosesan data, analisa data dan pengolahan data, dan evaluasi model yang telah dibuat[21]. Tahapan Penelitian dapat dilihat pada gambar 1 berikut.



Gambar. 1. Tahapan Penelitian

A. Identifikasi Masalah

Peneliti melakukan analisa permasalahan yang akan dibuatkan pemodelan topik pengambilan judul tugas akhir mahasiswa pada Program studi Sistem Informasi agar permodelan yang dibuat dapat lebih fokus dan spesifik. Dengan melakukan studi literatur, untuk menambah wawasan peneliti dalam membuat *topic modeling* yang akan dibuat, kemudian dilakukan wawancara dengan bagian Program studi Sistem Informasi untuk membahas lebih detail terkait permasalahan yang dialami dalam pengambilan judul tugas akhir agar model yang dibuat dapat menjawab masalah penelitian.

B. Data Collection

Studi kasus penelitian ini ditujukan pada Program studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana Salatiga. Program studi Sistem Informasi merupakan salah satu Program studi yang meluluskan banyak mahasiswa setiap tahunnya dan ada ratusan mahasiswa baru yang bergabung, maka dari itu terdapat banyak data judul tugas akhir yang berhasil didapatkan dari *repository* Perpustakaan O. Notohamidjojo - Universitas Kristen Satya Wacana dari tahun 2015 hingga 2021 yang akan menjadi *dataset*. *Dataset* berisikan judul, nama program studi, tahun terbit, nama penulis dan tautan *repository*. *Dataset* terdiri dari 2528 baris dan 5 kolom. *Dataset* ini diolah menggunakan layanan Google Colab. Berikut gambar *dataset* yang digunakan, bisa dilihat pada tabel 1.

TABEL I
 Preview *Dataset* Judul tugas akhir

Title	Researcher	Publish	Repository	Major
Perancangan dan Implementasi Aplikasi untuk Melakukan Transaksi Tarik Tunai pada Platform Android di PT Aristi Jasadata	Edon, Victoria Handayani	2018	http://repository.uksw.edu/handle/123456789/19088	T1 - Informatics Engineering
Perancangan dan Implementasi Sistem Informasi Akademik Berbasis Web (Studi Kasus : Universitas Kristen Tentena)	Sabintoe, Indriani Riski Yunita	2018-07-03	http://repository.uksw.edu/handle/123456789/15130	T1 - Informatics Engineering
Penerapan Virtual Router Redundancy Protocol pada Mikrotik Berbasis IPv6 untuk Meningkatkan Performansi Quality of Service	Iriansyah, Arief	2017	http://repository.uksw.edu/handle/123456789/13476	T1 - Informatics Engineering
Analisis dan Perancangan Sistem Informasi Inventor Menggunakan Web Service RESTful: Studi Kasus Toko Kosy Alat dan Bahan Kue Salatiga	Sutejo, Sri	2019	https://repository.uksw.edu/handle/123456789/19685	T1 - Informatics Engineering
Pengembangan Aplikasi Pembayaran SPP pada SMA Masehi Berbasis Web	Ananta, Andronico Pratama	2019	https://repository.uksw.edu/handle/123456789/19964	T1 - Information Systems

C. Analisis dan Eksplorasi Data

Dataset yang digunakan kemudian dianalisa agar dapat dipahami sehingga nantinya dapat diolah untuk mendapatkan atribut yang sesuai untuk dipergunakan dalam proses berikutnya. Dilakukan proses Exploratory Data Analysis (EDA) agar lebih mudah pemrosesan analisa *dataset* yang tersedia. Langkah ini juga digunakan untuk membantu pengoptimalan pengetahuan mengenai data, menampilkan variabel penting, mendeteksi data yang tidak diinginkan, dan anomali pada data. Sehingga model yang dibuat dapat menghasilkan kata-kata yang bermakna dan bermanfaat[13]. Berikut hasil proses *exploratory data analysis* pada *dataset* Judul tugas akhir dapat dilihat pada tabel 2.

TABEL II
 Hasil proses *Exploratory Data Analysis*

Title	Publish	Major
Pengembangan Aplikasi Pembayaran SPP pada SMA Masehi Berbasis Web	2019	T1 - Information Systems
Perencanaan Strategis Sistem Informasi di PT. PLN (Persero) Area Manado Menggunakan Metode Ward and Peppard	2018	T1 - Information Systems
Analysis of Information Technology Governance Using the COBIT 5 Framework: Case Study E-Legal Drafting Legal Section of the Regional Secretariat of Salatiga City	2019	T1 - Information Systems
Implementasi Curve Fitting pada Metode Indeks Musim sebagai Model Proyeksi untuk Optimasi Hasil Produksi: Studi Kasus Perumahan Curah Hujan Kabupaten Grobogan	2018	T1 - Information Systems
Sistem Informasi Promosi Berbasis Web pada Jimm's Bridal	2015	T1 - Information Systems

D. Data Pre - Processing

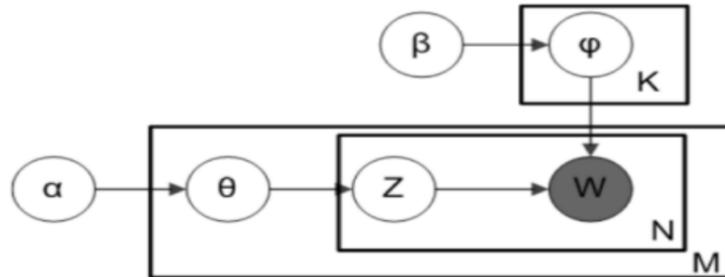
Untuk menyiapkan data agar dapat diolah lebih lanjut, dilakukan tahap *pre-processing*. *Dataset* akan diproses melalui beberapa tahapan agar data lebih bersih dan rapi sesuai dengan kebutuhan[22]. Peneliti hanya menggunakan data yang diperlukan saja dari *dataset* yang ada. Karena hanya ingin mengetahui pemetaan Judul tugas akhir untuk Mahasiswa Program studi Sistem Informasi, sedangkan di *dataset* memuat seluruh Judul tugas akhir di Fakultas Teknologi Informasi yang terdiri dari 8 Program studi.

Pada tahapan ini, dilakukan 4 tahapan yaitu *cleansing*, *stopword*, *case folding*, dan *tokenizing*. *Cleansing* dilakukan untuk memisahkan data berdasarkan Program studi yaitu hanya mengambil data Judul dari Program studi Sistem Informasi saja, kemudian kolom yang berisikan format tanggal yang kurang terstruktur akan diekstrak atau diambil *value* tahunnya, meninggalkan kolom tautan *repository* perpustakaan dan nama *author*. *Stopword* dilakukan pembentian terhadap kata umum yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna, misal kata depan, kata sambung, dan kata ganti. Atau sederhananya, melakukan removal terhadap kata yang tidak diinginkan[14]. Tahapan *case folding* mengubah semua karakter pada *dataset* yang bentuknya masih campuran huruf besar dan kecil menjadi *lowercase*[15]. Pada tahap *Tokenizing* dilakukan pemecahan kalimat menjadi potongan karakter kata-kata yang membentuknya[16]Sederhananya yaitu pembagian teks menjadi kata atau frase yang disebut token[17].

E. Pembuatan Topic Modeling

Topic modeling adalah sebuah teknik untuk menemukan topik atau tema utama dari sejumlah besar dokumen tidak terstruktur yang berbentuk teks yang dapat digunakan untuk membuat suatu kesimpulan, termasuk topik tersembunyi yang menghasilkan kumpulan frasa dan kata yang menurut mereka terkait [21][23]. Topik berisi sekelompok kata-kata yang sering muncul bersamaan, dan pemodelan topik dapat menghubungkan kata-kata yang memiliki arti serupa dan dapat membedakan antara penggunaan kata dengan banyak makna (Alghamdi & Alfalqi, 2015).Sederhananya pemodelan topik adalah teknik penambangan teks untuk menemukan topik (Blei, 2012). Tujuan *topic modeling* yaitu menentukan topik secara otomatis dari sekumpulan dokumen dengan pengkodean[6][24].

Pada penelitian ini digunakan teknik Latent Dirichlet Allocation (LDA) yang dikembangkan pada tahun 2003 oleh peneliti David Blei, Andrew Ng dan Michael Jordan. Merupakan salah satu metode paling populer dalam pemodelan topik. LDA dapat digunakan untuk meringkas, klusterisasi, menghubungkan dan memproses data yang sangat besar karena LDA mampu menampilkan daftar topik dan membandingkan besar kemunculan topik[27]. Dalam proses penerapan LDA, digunakan *library* gensim sebagai *document indexing* dan *similarity retrieval* terhadap data yang sudah melewati tahap data *pre-processing* menggunakan kode program Python yang akan diimplementasikan melalui layanan Google Colab. Notasi dalam LDA dapat dilihat pada gambar 2.



Gambar. 2. Notasi LDA

Simbol β merupakan dirichlet parameter atas distribusi kata terhadap topik, ϕ merupakan distribusi kata terhadap topik dalam corpus, K merupakan kumpulan topik, W merupakan kata, N merupakan kumpulan kata, M merupakan dokumen, Z merupakan topik *index assignment*, θ merupakan dokumen, dan α merupakan dirichlet parameter atas distribusi topik terhadap dokumen.

LDA dirumuskan sebagai berikut:

$$P(W, Z, \theta, \phi; \alpha, \beta) = \prod_{j=1}^M P(\theta_j; \alpha) \prod_{i=1}^K P(\phi_i; \beta) \prod_{t=1}^N P(Z_{j,t} | \theta_j) P(W_{j,t} | \phi_{j,t})$$

Modeling dengan Latent Dirichlet Allocation (LDA)

Library yang digunakan :

- Pandas adalah *library* yang digunakan untuk keperluan analisis data, salah satunya untuk mengecek data.
- NLTK, adalah *library* untuk pengolahan data untuk melakukan *tokenize*.
- Gensim merupakan *library* dari Python yang dikhususkan dalam *Natural Language Processing*, salah satunya *Topic Modeling* dalam *corpus* dengan jumlah yang besar.
- *Regular expression library* untuk mengedit data.
- *CoherenceModel library* untuk mengevaluasi model.
- *PyLDAvis library* untuk visualisasi hasil model yang berbasis web interaktif.
- *Wordcloud library* untuk menampilkan daftar kata masing-masing *cluster*.

F. Evaluasi

Pada tahap ini model yang sudah dibuat akan melewati proses evaluasi untuk mendapatkan performa yang lebih optimum dari *Text modelling*. Evaluasi dilakukan dengan metode *hyperparameter tuning* untuk menemukan nilai parameter yang dapat menghasilkan model dengan performa yang paling baik untuk *learning rate* pada *neural network*. Untuk mendapatkan parameter yang terbaik bagi model dilakukan uji coba dengan melakukan optimasi pada nilai *coherence* yaitu dilakukan peningkatan nilai *coherence* dari nilai *coherence* sebelumnya dan menemukan nilai masing-masing parameter lain seperti parameter *alpha*, *beta*, dan jumlah *topic* secara otomatis dijalankan sebanyak 540 kali cukup dalam 1 kali *running code* yang membantu menemukan nilai untuk parameter *alpha*, *beta*, dan *coherence score* yang sudah ditingkatkan, serta jumlah *topic* dalam bentuk file CSV yang dihasilkan dari proses *tuning*. *Tuning* cukup memakan waktu yang lumayan lama karena pada proses tersebut dilakukan pembuatan kombinasi masing-masing parameter dan akan menghasilkan ratusan uji coba. Sehingga akan didapat nilai parameter-

parameter yang diusulkan untuk model prediktif pada *dataset*[25]. Parameter yang digunakan yaitu *alpha* adalah parameter yang mengontrol pre-distribusi bobot topik di setiap dokumen, sedangkan *beta* adalah parameter untuk pre-distribusi bobot kata di setiap topik dan *coherence score* merupakan alat ukur kesamaan *semantic* antara kata-kata yang mendapatkan skor tertinggi dalam topik[6][21]. Sederhananya hasil pengujian *hyperparameter tuning* dapat digunakan untuk mengatur kombinasi yang tepat untuk *final* model sehingga memungkinkan model untuk memaksimalkan kinerjanya[21]. Pada *final* model, akan diisi oleh parameter yang sudah didapat dari hasil tuning, cara mendapatkannya yaitu dengan mengunduh file hasil *tuning*. Dari file tersebut kemudian dicari nilai *coherence* yang paling maksimum hasil dari *improved* pada proses *tuning*[6]. Hasil *coherence* sebelum *tuning* yaitu 0,53448 menjadi sebesar 0,617789 dengan 10 jumlah *topic*, 0,91 *alpha* dan 0,01 *beta*.

III. HASIL DAN PEMBAHASAN

Pada bagian ini akan dijelaskan mengenai detail hasil dan pembahasan dari performa model yang telah dibuat dengan metode yang digunakan untuk penelitian ini yaitu Latent Dirichlet Allocation (LDA). Mulai dari pemrosesan data seperti : *Stopword*, *Case folding*, dan *Tokenizing*. Pada tahap *Stopword*, dilakukan pembentian terhadap kata umum yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna, misal kata depan, kata sambung, dan kata ganti. Atau sederhananya, melakukan removal terhadap kata yang tidak diinginkan[14]. Hasil dapat dilihat pada tabel 3.

TABEL III
 Hasil proses Stopword

Sebelum	Sesudah
Perencanaan Strategis Sistem Informasi di PT. PLN (Persero) Area Manado Menggunakan Metode Ward and Peppard	'Perencanaan' 'Strategis' 'Sistem' 'Informasi' 'PLN' ('Persero') 'Area' 'Manado' Menggunakan 'Metode' 'Ward' 'and' 'Peppard'

Pada tahap *case folding* mengubah semua karakter pada *dataset* yang bentuknya masih campuran huruf besar dan kecil menjadi *lowercase* [15] Hasil dapat dilihat pada tabel 4.

TABEL IV
 Hasil proses Case Folding

Sebelum	Sesudah
Perencanaan Strategis Sistem Informasi di PT. PLN (Persero) Area Manado Menggunakan Metode Ward and Peppard	perencanaan strategis sistem informasi di pt pln (persero) area manado menggunakan metode ward and peppard

Pada tahap *Tokenizing* dilakukan pemecahan kalimat menjadi potongan karakter kata-kata yang membentuknya[16]Sederhananya yaitu pembagian teks menjadi kata atau frase yang disebut token[17]. Hasil *tokenizing* dapat dilihat pada tabel 5.

TABEL V
 Hasil proses *Tokenizing*

Sebelum	Sesudah
Perencanaan Strategis Sistem Informasi di PT. PLN (Persero) Area Manado Menggunakan Metode Ward and Peppard	'Perencanaan' 'Strategis' 'Sistem' 'Informasi' 'di' 'PT' 'PLN'' ('Persero')' 'Area' 'Manado'' Menggunakan 'Metode' 'Ward' 'and' 'Peppard'

Kemudian dilakukan pembuatan *clustering* dengan menggunakan LDA dan Gensim menggunakan *dataset* yang sudah bersih yaitu tugas akhir mahasiswa Sistem Informasi yang sudah disiapkan dan sudah melalui tahap *pre-processing* untuk membuat *topic cluster*. *Pre-processing* data yang dimaksud adalah membuang kata yang tidak perlu dan mengambil data yang diperlukan saja[18]. Selanjutnya dilakukan *import CoherenceModel* untuk menghasilkan *coherence score*. Dan didapatkan *coherence score* sebesar 0,53448.

```
# Build LDA model
lda_model = gensim.models.LdaMulticore(corpus=corpus,
                                       id2word=id2word,
                                       num_topics=10,
                                       random_state=100,
                                       chunksize=100,
                                       passes=10,
                                       per_word_topics=True)
```

Gambar. 3. Pembuatan model LDA menggunakan gensim

Setelah melewati pembuatan model dengan LDA dan Gensim untuk membuat *topic cluster*, kemudian dilakukan *hyperparameter tuning* untuk evaluasi model. Hasil evaluasi mendapatkan jumlah *cluster*, parameter *alpha* dan *beta* serta nilai *coherence score* untuk dimasukkan ke dalam *final* model. Diambil nilai *coherence score* paling besar agar kesamaan antara dokumen menjadi lebih mirip dalam topik yang dikandungnya karena nilai maksimum adalah hasil *improved* pada proses *tuning*[6][21]. Hasilnya akan diamati agar dapat diketahui makna dari suatu kata tertentu[19]. Dari hasil *tuning* diperoleh hasil yang optimum yang akan digunakan untuk *final* model sebagai berikut: 10 jumlah *topic*, 0,91 *alpha* dan 0,01 *beta* dan *coherence score* sebesar 0,617789.

```
num_topics = 10

lda_model = gensim.models.LdaMulticore(corpus=corpus,
                                       id2word=id2word,
                                       num_topics=num_topics,
                                       random_state=100,
                                       chunksize=100,
                                       passes=10,
                                       alpha=0.91,
                                       eta=0.01)
```

Gambar. 4. *Final* model

Karena jumlah *cluster* yang ingin dihasilkan adalah 10, maka dihasilkan 10 *topic cluster*. Dari 10 *cluster* yang sudah didapat, dilakukan analisa terhadap kata-kata yang muncul di tiap *cluster* dan terdapat 4 *cluster* yang akan digunakan karena dapat dianalisa dengan jelas. Berikut judul tugas akhir dari masing-masing *cluster* dapat dilihat pada tabel 6, 7, 8, dan 9.

TABEL VI
 Judul tugas akhir pada *cluster* 1

<i>Cluster</i> ke 1 (<i>Topic</i> 1)
<ol style="list-style-type: none"> 1. evaluasi tingkat penerimaan dan penggunaan aplikasi simda menggunakan technology acceptance model (tam) pada kantor bupati maluku tengah. 2. analisis penerimaan pengguna terhadap aplikasi salatiga mobile library menggunakan technology acceptance model (tam): studi kasus perpustakaan dan arsip daerah kota salatiga. 3. evaluasi dan analisis penerimaan sistem evaluasi dosen pada universitas kristen satya wacana menggunakan technology acceptance model (tam). 4. analisis pemanfaatan aplikasi traveloka menggunakan technology acceptance model (tam) dan usability pada pengguna aplikasi traveloka di uksw. 5. analisis pemanfaatan teknologi informasi dengan menggunakan metode technology acceptance model (tam) pada sma negeri 1 seram barat.

TABEL VII
 Judul tugas akhir pada *cluster* 2

<i>Cluster</i> ke 2 (<i>Topic</i> 2)
<ol style="list-style-type: none"> 1. analisis pemanfaatan teknologi informasi menggunakan framework cobit 4.1 domain plan and organise pada pt. batamtex, ungaran. 2. evaluasi manajemen risiko teknologi informasi pada perusahaan daerah air minum kota salatiga menggunakan framework cobit 5.0. 3. analisa kinerja si ti pada badan pelayanan perizinan terpadu dan penanaman modal kota salatiga menggunakan framework cobit 5. 4. evaluasi kinerja teknologi informasi menggunakan framework cobit pada perusahaan distribusi barang: studi kasus pt.intra-sari raya,semarang. 5. evaluasi tata kelola teknologi informasi menggunakan framework cobit 5 domain edm pada yayasan bina darma salatiga.

TABEL VIII
 Judul tugas akhir pada *cluster* 3

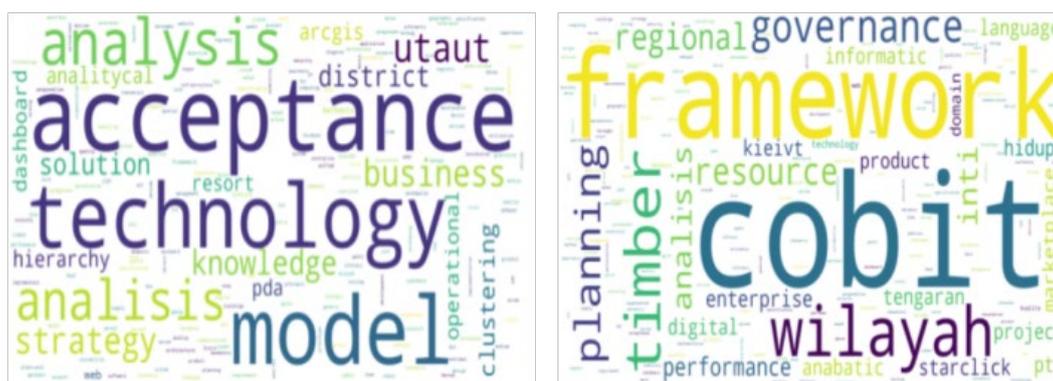
<i>Cluster</i> ke 3 (<i>Topic</i> 3)
<ol style="list-style-type: none"> 1. evaluasi tata kelola teknologi informasi pt.asurnasi sinarmas menggunakan framework cobit 5 domain dss (deliver,service, and support) 2. an analysis of information technology governance in department of communication and informatics of salatiga using cobit 5 framework dss domain 3. analisis kualitas layanan teknologi informasi menggunakan framework itil v3 domain service operation: studi kasus divisi software development 6 indomaret group 4. analisis kualitas layanan teknologi informasi menggunakan framework information technology infrastructure library version 3 domain service operation pada pt pln rayon manado utara 5. analisis tata kelola teknologi informasi e-ktip menggunakan framework cobit 5 domain dss (deliver, service, support): studi kasus dinas kependudukan dan pencatatan sipil kabupaten landak

TABEL IX
 Judul tugas akhir pada *cluster* 4

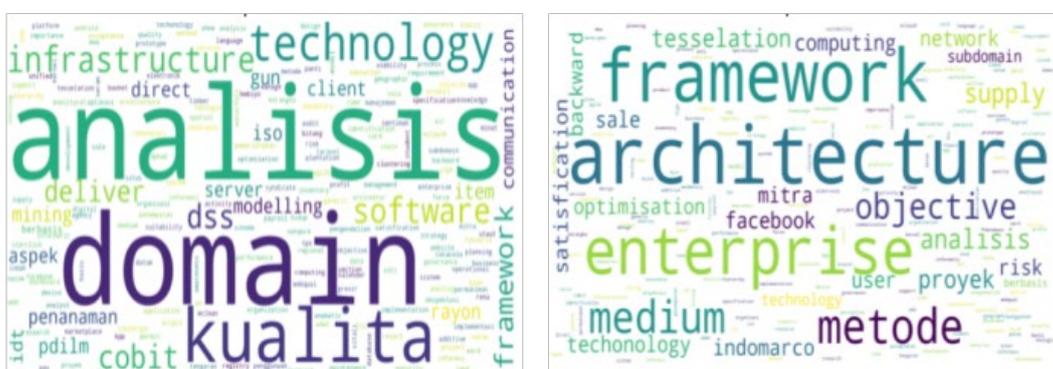
Cluster ke 4 (*Topic* 4)

1. perencanaan arsitektur enterprise menggunakan framework togaf (studi kasus: ltc-uksw).
2. perencanaan strategis sistem informasi menggunakan kerangka the open group architecture framework (togaf) pada bpjs kesehatan surakarta.
3. perencanaan strategis sistem informasi menggunakan the open group architecture framework (togaf) (studi kasus : gki masaran).
4. perancangan enterprise architecture (ea) pada perguruan tinggi menggunakan the open group architecture framework (togaf) (studi kasus: universitas kristen satya wacana salatiga).
5. perencanaan strategis sistem informasi menggunakan kerangka the open group architecture framework (togaf): studi kasus rs.banjar patroman.

Untuk membaca data teks yang paling sering muncul pada masing-masing *cluster* dengan mudah, dibuat *wordcloud* untuk setiap topik yang sudah dipilih. *Topic* 1 mengandung kata yang mengarah ke penggunaan *Technology Acceptance Model*, *topic* 2 mengandung kata yang mengarah ke penggunaan *Framework Cobit* dalam Audit. *Topic* 1 dan 2 dapat dilihat pada gambar 3. Kemudian *topic* 3 mengandung kata yang mengarah ke *Framework Cobit* sama seperti pada *topic* 2, *topic* 4 mengandung kata yang mengarah ke penerapan Arsitektur *Enterprise*. *Wordcloud* *topic* 3 dan 4 dapat dilihat pada gambar 4. Dan Analisa setiap *cluster* juga dapat dilihat pada tabel 10.



Gambar. 5. *Wordcloud* *topic* 1 dan 2

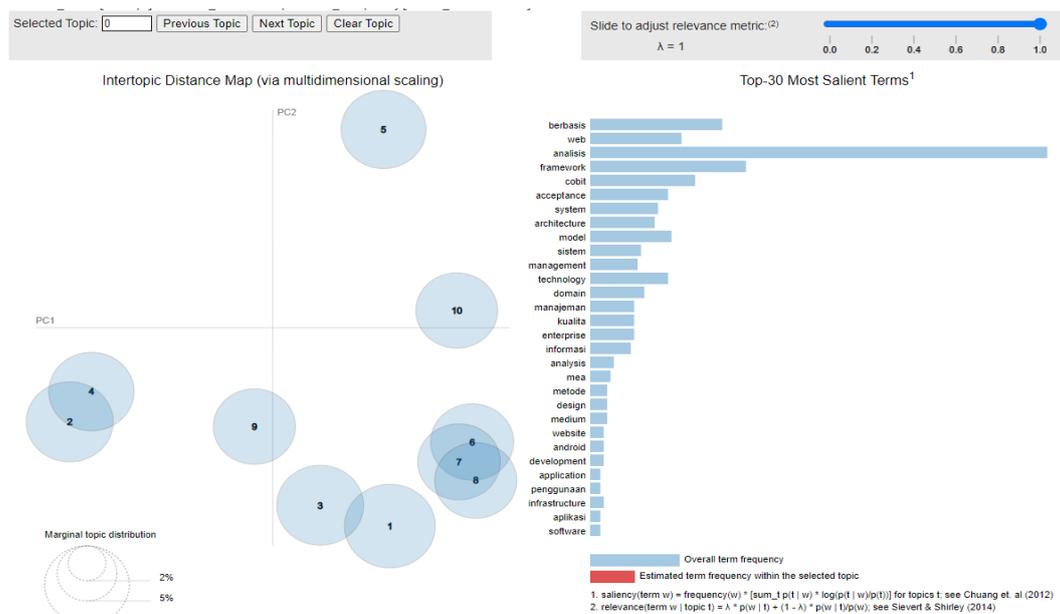


Gambar. 6. *Wordcloud* topik 3 dan 4

TABEL X
 Hasil Analisa tiap *topic*

<i>Topic</i>	Analisa
1	Analisis dan Evaluasi Technology Acceptance Model (TAM) merupakan model yang digunakan dalam memprediksi persepsi terhadap penggunaan dan penerimaan teknologi informasi yang berkaitan dalam penggunaan sistem informasi setelah dilakukan penerapan sebuah sistem informasi dengan tujuan agar dapat diketahui tingkat manfaat dan pengaruh terhadap pengguna. TAM berhubungan dengan matakuliah Perencanaan Strategis Sistem Informasi atau Perancangan Sistem Informasi.
2	Analisis dan Evaluasi Framework Cobit sangat berhubungan dengan kegiatan Audit Tata Kelola Teknologi Informasi untuk mendukung perusahaan/organisasi untuk mengelola Teknologi Informasi agar dapat bermanfaat sesuai tujuan bisnis. Cobit biasanya muncul di matakuliah Audit Tata Kelola TI dan Management Risiko Teknologi Informasi.
3	Analisis dengan domain dss masih berhubungan dengan Framework Cobit dalam memastikan bahwa seluruh solusi teknologi informasi yang sudah diimplementasikan berhasil melayani dan mendukung pengguna sesuai dengan harapan.
4	Perencanaan Arsitektur Enterprise berhubungan dengan Framework Togaf yang digunakan untuk mengembangkan Software. Arsitektur Enterprise digunakan untuk merancang sebuah sistem untuk mencapai tujuan bisnis dengan penerapan Teknologi. Arsitektur Enterprise merupakan salah satu matakuliah yang sudah diberikan di awal memasuki pertengahan kuliah di Prodi Sistem Informasi.

Setelah dibuat pemodelan topik yang sudah mencapai performa yang optimum, dilakukan visualisasi dalam bentuk pyLDAvis yang dapat menampilkan masing-masing *topic* dan setiap kata yang sering muncul di *dataset* yang sudah diolah[20]. Masing-masing *bubble* atau lingkaran mewakili tiap *topic*. Semakin jauh *bubble* satu dengan yang lainnya atau tidak saling tumpang tindih maka semakin mudah untuk menginterpretasikan setiap *topic*[26]. *Topic* 1 dan 3 memiliki hubungan yang erat dan keterkaitan atau disebut *intertopic distance map*, ditunjukkan dari irisan yang menghubungkan kedua *topic* yang artinya memiliki distribusi probabilitas kata-kata yang mirip. Kemudian *topic* 2 dan 4 juga memiliki hubungan *intertopic distance map* sama seperti *topic* 1 dan 3. Pada *topic-term distribution* menggambarkan kontribusi kata-kata terhadap topik-topik yang tersembunyi, dapat dilihat pada batang biru merupakan kata yang paling sering muncul dari keseluruhan *topic*, kata “analisis” memiliki frekuensi yang paling besar. Untuk melihat top 30 kata tersembunyi, dapat dilihat pada gambar 7.



Gambar. 7. Visualisasi dengan pyLDAvis

IV. KESIMPULAN

Berdasarkan analisis terhadap hasil dari pemodelan menggunakan LDA dengan performa model yang telah dilakukan *tuning* parameter yang menghasilkan nilai *coherence* dari 0,53448 menjadi 0,617789. Dapat disimpulkan bahwa metode LDA mampu mereduksi setiap kata dengan baik sehingga menghasilkan 4 *cluster* bertema Tata Kelola dari topik 1,2,3 dan 4 yaitu mengenai *Technology Acceptance Model*, *Framework Cobit*, dan *Arsitektur Enterprise* yang dapat menjadi referensi dan landasan pengambilan judul tugas akhir untuk mahasiswa sebagai dasar pengetahuan minat judul tugas akhir di program studi. Selain itu, dengan diketahui *cluster* minat tugas akhir mahasiswa dapat memberikan pengetahuan dan kontribusi bagi program studi sebagai landasan penyesuaian judul tugas akhir mahasiswa, konsentrasi serta pengembangan kurikulum di program studi. Dengan demikian dapat disimpulkan juga bahwa pemodelan topik menggunakan metode LDA sesuai dengan hasil penelitian terdahulu yaitu berhasil memodelkan topik dengan baik. Peneliti selanjutnya dapat melakukan penelitian serupa dengan studi kasus yang sama namun dengan konfigurasi *final* model yang memiliki nilai *alpha*, *beta*, dan *coherence score* yang lebih rendah dengan uji coba sebanyak mungkin serta mengkombinasikan ketiga parameter dengan nilai rendah dan tinggi. Sehingga menambah penelitian yang membahas terkait evaluasi *topic modeling* lebih detail dan mengarah ke eksperimen nilai *coherence* dalam pembuatan *topic modeling* yang optimal.

V. DAFTAR PUSTAKA

- [1] Zain, M. I., Radiusman, R., Syazali, M., Hasnawati, H., & Amrullah, L. W. Z. (2021). IDENTIFIKASI KESULITAN MAHASISWA DALAM PENYUSUNAN SKRIPSI PRODI PGSD UNIVERSITAS MATARAM. *Tunjuk Ajar: Jurnal Penelitian Ilmu Pendidikan*, 4(1), 73. <https://doi.org/10.31258/jta.v4i1.73-85>
- [2] Kurniawan, H., & Bondowoso, W. B. (2019). SISTEM INFORMASI TERINTEGRASI TUGAS AKHIR/SKRIPSI BERBASIS WEB (Studi Kasus: Jurusan Sistem Informasi Institut Informatika dan Bisnis Darmajaya). In *Jurnal Sistem Informasi dan Manajemen Basis Data (SIMADA)* (Vol. 02, Issue 02).
- [3] Subekti, P., Andini, T. D., & Islamiyah, M. (2022). Sistem Penentuan Konsentrasi Jurusan Bagi Mahasiswa Informatika Menggunakan Metode K-Means Di Institut Asia Malang. *Jurnal Manajemen Informatika (JAMIKA)*, 12(1), 25–39. <https://doi.org/10.34010/jamika.v12i1.6452>
- [4] Nurlayli, A., & Nasichuddin, M. A. (2019). Topic Modeling Penelitian Dosen JPTEI UNY pada Google Scholar Menggunakan Latent Dirichlet Allocation. 4(2), 154–161. <https://doi.org/10.21831/elinvo.v4i2>
- [5] Sahria, Y., Isnaini Febriarini, N., Dwi Oktavianti, P., Teknologi Yogyakarta, U., Studi Sarjana Keperawatan Al Islam Yogyakarta, P., Siliwangi Jl Ring Road Utara, J., Lor, J., Mlati, K., & Sleman, K. (n.d.). *Yoga Sahria, Pemodelan Topik Penelitian Bidang Keperawatan Indonesia pada Repository Jurnal Sinta Menggunakan Metode Topic modeling LDA (Latent Dirichlet Allocation) PEMODELEN TOPIK PENELITIAN BIDANG KEPERAWATAN INDONESIA PADA REPOSITORY JURNAL SINTA MENGGUNAKAN METODE TOPIC MODELING LDA (LATENT DIRICHLET ALLOCATION)*.
- [6] Storopoli, J. E. (n.d.). Perspectives TOPIC MODELING: HOW AND WHY TO USE IN MANAGEMENT RESEARCH. <https://doi.org/10.1177/2329488>

- [7] Westerlund, M., Leminen, S., & Rajahonka, M. (2018). A Topic modeling Analysis of Living Labs Research. In *Technology Innovation Management Review* (Vol. 8, Issue 7).
- [8] Mailoa, E., Damayanti, W., & Christopher, N. (2021). The Use of Text Mining to Investigate Undergraduate Research Topic Clusters from 2015 to 2019. *2021 2nd International Conference on Innovative and Creative Information Technology, ICITech 2021*, 44–49.
- [9] Iffan Alfanzar, A., & Sudanawati Rozas, I. (n.d.). TOPIC MODELING SKRIPSI MENGGUNAKAN METODE LATENT DIRICLHET ALLOCATION. *Sistem Informasi* |, 7(1), 7–13.
- [10] Novarian, N., Khomsah, S., & Arifa, A. B. (2023). *LEDGER: Journal Informatic and Information Technology Topic Modeling tugas akhir Mahasiswa Fakultas Informatika Institut Teknologi Telkom Purwokerto Menggunakan Metode Latent Dirichlet Allocation* (Vol. 2, Issue 1).
- [11] Kherwa, P., & Bansal, P. (2020). Topic Modeling: A Comprehensive Review. *EAI Endorsed Transactions on Scalable Information Systems*, 7(24), 1–16. <https://doi.org/10.4108/eai.13-7-2018.159623>
- [12] Naury, C., Fudholi, D. H., & Hidayatullah, A. F. (2021). Topic modeling pada Sentimen Terhadap Headline Berita Online Berbahasa Indonesia Menggunakan LDA dan LSTM. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 5(1), 24. <https://doi.org/10.30865/mib.v5i1.2556>
- [13] Wayan, N., Suaryani, N., Putri, T., Rifaldi, R., & Giri, W. (2020). ANALISIS TOPIC MODELING UNTUK MENGIDENTIFIKASI TOPIK PEMBICARAAN PADA MEDIA SOSIAL BANK BNI Jurnal Mitra Manajemen (JMM Online). 4(6), 1002–1013.
- [14] Sihombing, D. O. (2022). Implementasi Natural Language Processing (NLP) dan Algoritma Cosine Similarity dalam Penilaian Ujian Esai Otomatis. 4, 396–406. <https://doi.org/10.30865/json.v4i2.5374>
- [15] Kurniasari, T. K., Maharani, W., & Husen, J. H. (n.d.). *Analisis Media Sosial Twitter untuk Mengetahui Pengguna Berpengaruh pada Portal Berita di Indonesia dengan Metode TSM (Topic-based Social Influence Measurement)*.
- [16] N. Hafidz and D. Y. Liliana, “Klasifikasi Sentimen pada Twitter Terhadap WHO,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 2, pp. 213–219, 2021, doi: <https://doi.org/10.29207/resti.v5i2.2960>.
- [17] Fitri, H., Widyawan, W., & Soesanti, I. (2021). Topic Modeling in the News Document on Sustainable Development Goals. *IJITEE (International Journal of Information Technology and Electrical Engineering)*, 5(3), 82. <https://doi.org/10.22146/ijitee.67467>
- [18] Keputusan Dirjen Penguatan Riset dan Pengembangan Ristek Dikti, S., Sahria, Y., & Hatta Fudholi, D. (2017). Terakreditasi SINTA Peringkat 2 Analisis Topik Penelitian Kesehatan di Indonesia Menggunakan Metode Topic Modeling LDA (Latent Dirichlet Allocation). *Masa Berlaku Mulai*, 1(3), 336–344.
- [19] Ridhwanulah, D., & Fudholi, D. H. (2022). Pemodelan Topik pada Cuitan tentang Penyakit Tropis di Indonesia dengan Metode Latent Dirichlet Allocation. *Jurnal Ilmiah SINUS*, 20(1), 11. <https://doi.org/10.30646/sinus.v20i1.589>
- [20] Dirjen, S. K., Riset, P., Pengembangan, D., Dikti, R., & Susanto, I. K. (2018). *Terakreditasi SINTA Peringkat 4 Analisis Sentimen dan Topic modeling Pada Pembelajaran Online di Indonesia Melalui Twitter* (Vol. 3, Issue 1).
- [21] Gupta, R. K., Agarwalla, R., Naik, B. H., Evuri, J. R., Thapa, A., & Singh, T. D. (2022). Prediction of research trends using LDA based topic modeling. *Global Transitions Proceedings*, 3(1), 298–304. <https://doi.org/10.1016/j.gltp.2022.03.015>
- [22] Aditya, B. R., No, J. T., Buah, T., & Bandung, B. (2015). Penggunaan Web Crawler Untuk Menghimpun Tweets dengan Metode Pre-Processing Text Mining. *Jurnal Infotel*, 7(2).
- [23] Jagannathan, M., Roy, D., & Delhi, V. S. K. (2022). Application of NLP-based topic modeling to analyse unstructured text data in annual reports of construction contracting companies. *CSI Transactions on ICT*, 10(2), 97–106. <https://doi.org/10.1007/s40012-022-00355-w>
- [24] Dheanis, K., Salsabila, A., Trianasari, N., Artikel, R., Kunci, K., & Konsumen, P. (2021). Jurnal Teknologi dan Manajemen Informatika Analisis Persepsi Produk Kosmetik Menggunakan Metode Sentiment Analysis Dan Topic Modeling (Studi Kasus: Laneige Water Sleeping Mask) Info Artikel ABSTRAK. 7(1). <http://http://jurnal.unmer.ac.id/index.php/jtmi>
- [25] Iqbal, M., Mahmud Nawawi, H., Rangga Ramadhan Saelan, M., Sony Maulana, M., & Mustopa, A. (2023). OPTIMASI HYPERPARAMETER MULTILAYER PERCEPTRON UNTUK PREDIKSI DAYA BELI MOBIL. *Jurnal Manajemen Informatika & Sistem Informasi (MISI)*, 6(1). <https://doi.org/10.36595/misi.v5i2>
- [26] Amaradiena, K., & Widarmanti, T. (2023). SEIKO : Journal of Management & Business LDA-Topic Modeling: Menggunakan Ulasan Pengguna Untuk Meningkatkan User Experience (Studi pada PeduliLindungi). *SEIKO : Journal of Management & Business*, 6(1), 943–953. <https://doi.org/10.37531/sejaman.v6i1.3802>
- [27] Rachman, F. F., & Pramana, S. (2020). Analisis Sentimen Pro dan Kontra Masyarakat Indonesia tentang Vaksin COVID-19 pada Media Sosial Twitter. In *Health Information Management Journal ISSN* (Vol. 8, Issue 2).