



Analisis Topik Penelitian Pendidikan Matematika Di Indonesia Dengan Menggunakan Metode *Latent Dirichlet Allocation* (LDA)

Beni Junedi ^{1*}, Agung Budi Susanto ², Sajarwo Anggai ³

^{1,2,3})Program Studi Teknik Informatika S-2, Universitas Pamulang

Email: ¹benijunedi07@gmail.com, ²dosen02680@unpam.ac.id, ³sajarwo@mail.com

ABSTRACT

On the research topic of Mathematics Education readers or researchers still have difficulty identifying research topics in the field of Mathematics Education. This is because there is no system or model that can be seen or used in determining research topics. Besides that, there is no automation of the research direction of Mathematics Education in Indonesia using topic modeling, so it is necessary to conduct a study or research on this. In research, the most important thing is the trend of research that is currently developing so that it can determine the novelty of the studies that have been done before. While there is no system used to determine trends and state of the art from research in the field of Mathematics Education. The aim of the research is to find out an overview of the research topics in Mathematics Education in Indonesia in 2020-2023 and to find out the implementation of modeling research topics in Mathematics Education in Indonesia using the Latent Dirichlet Allocation (LDA) method for 2020-2023. The research design consisted of literature study, data collection, data pre-processing: tokenization, case folding, stopword removal, and stemming, topic analysis with LDA, evaluation of the LDA method, and conclusions. Analysis of Topic Modeling with Latent Dirichlet Allocation using packages used from python including the Gensim and pyLDAvis packages. Based on the coherence score, the best number of topics (K) = 18, with a coherence score = 0.426 (the highest), it can be concluded that the number of topics produced is 18 topics.

Keywords: Topic Modeling, Latent Dirichlet Allocation (LDA), Scientific publication abstracts, Mathematics Education

ABSTRAK

Pada topik penelitian Pendidikan Matematika pembaca atau peneliti masih mengalami kesulitan dalam mengidentifikasi topik penelitian di bidang Pendidikan Matematika. Hal ini disebabkan karena belum adanya sistem atau model yang dapat dilihat atau digunakan dalam penentuan topik penelitian. Disamping itu belum adanya otomatisasi arah penelitian Pendidikan Matematika di Indonesia dengan menggunakan pemodelan topik, sehingga perlu dilakukan kajian atau penelitian tentang ini. Dalam penelitian suatu hal yang terpenting adalah trend penelitian yang sedang berkembang sehingga dapat menentukan novelty dari penelitian-penelitian yang sudah dilakukan sebelumnya. Sedangkan belum adanya sistem yang digunakan untuk menentukan trend dan *state of the art* dari penelitian di bidang Pendidikan Matematika. Tujuan dari penelitian yaitu mengetahui gambaran umum dari topik penelitian Pendidikan Matematika di Indonesia tahun 2020-2023 dan mengetahui implementasi pemodelan topik penelitian Pendidikan Matematika di Indonesia dengan menggunakan metode *Latent Dirichlet Allocation* (LDA) Tahun 2020-2023. Rancangan penelitian terdiri dari studi literatur, pengumpulan data, pre-prosesing data: *tokenization*, *case folding*, *stopword removal*, dan *stemming*, analisis topik dengan LDA, evaluasi Metode LDA, dan kesimpulan. Analisis *Topic Modeling* dengan *Latent Dirichlet Allocation* menggunakan packages yang digunakan dari python diantaranya yaitu packages Gensim dan pyLDAvis. Berdasarkan *coherence score* diperoleh jumlah topik terbaik (K) = 18, dengan *coherence score* = 0.426 (tertinggi), dapat disimpulkan bahwa jumlah topik yang dihasilkan yaitu 18 topik.

Kata kunci: Pemodelan Topik, *Latent Dirichlet Allocation* (LDA), Abstrak Publikasi Ilmiah, Pendidikan Matematika

1. PENDAHULUAN

Text mining merupakan pengolahan data dalam bentuk teks. Pengolahan data teks ini dilakukan dari yang tidak terstruktur (*unstructured data*) kemudian diolah dengan menggunakan *text mining* [1]. *Text mining* menganalisis text yang tidak terstruktur yang saling terkait satu sama lain serta berhubungan dengan prinsip yang sama [2]. Hasil dari proses *text mining* berupa pemahaman baru yang yang belum diketahui sebelumnya. *Text mining* juga diartikan sebagai proses mengeksplorasi dan menganalisis sejumlah besar data teks tidak terstruktur dengan bantuan perangkat lunak yang dapat mengidentifikasi konsep, pola, topik, kata kunci, dan atribut lain dalam data [3]. *Text mining* adalah proses mengubah teks yang tidak terstruktur menjadi format terstruktur untuk mengidentifikasi pola yang bermakna dan wawasan baru [4].

Text mining bertujuan untuk untuk memperoleh informasi dari beberapa dokumen [5]. Sedangkan proses *text mining* bertujuan untuk memproses data tidak terstruktur untuk mengekstrak pola sehingga dapat ditindaklanjuti untuk pengambilan keputusan yang lebih baik. Tujuan pengolahan teks adalah mengetahui dan mengekstrak informasi yang berguna dari sumber data dengan identifikasi dan eksplorasi pola menarik [6].

Beberapa penerapan *text mining* yaitu ekstraksi informasi (*information extraction*), pelacakan topik (*topic tracking*), Perangkuman (*Summarization*), Kategorisasi (*Categorization*), Penggugusan (*Clustering*), Penautan Konsep (*Concept Linking*), dan Penjawaban Pertanyaan (*Question Answering*) [7]. *Information extraction* (ekstraksi informasi) merupakan identifikasi terhadap hubungan dan frase-frase kunci dalam text dengan mencari urutan yang sudah ditetapkan dalam text menggunakan pencocokan pola [8]. *Topic tracking* (pelacakan topik) merupakan *text mining* pada profil user dan berbagai dokumen yang dilihat user, *text mining* bisa memprediksi dokumen-dokumen lain yang menjadi perhatian/minat user tersebut [9]. *Summarization* (peringkasan) merupakan meringkas suatu dokumen untuk menghemat waktu dari sisi pembaca [10]. *Clustering* merupakan pengelompokkan dokumen-dokumen yang mirip tanpa memiliki kategori yang sudah ditetapkan sebelumnya [11]. *Concept linking* merupakan menghubungkan berbagai dokumen terkait dengan mengidentifikasi konsep yang digunakan bersama dan dengan demikian membantu para user untuk menemukan informasi yang tidak ditemukan dengan menggunakan metode-metode pencarian tradisional [12]. *Question answering*

merupakan menemukan jawaban terbaik pada pertanyaan yang diberikan melalui pencocokan pola berbasis *knowledge* [13].

Beberapa penerapan text mining salah satunya tentang *Clustering*. *Clustering* merupakan salah satu teknik dari algoritma *machine learning* yaitu *unsupervised learning*, dimana membagi dokumen pada sifat yang sama kedalam beberapa kelompok kecil untuk dikelompokkan [14]. Tujuan akhir *clustering* adalah untuk menentukan kelompok dalam sekumpulan data yang tidak berlabel, dimana *clustering* merupakan metode *unsupervised* [15]. *Clustering* dapat dilakukan pada topik penelitian yang ada pada jurnal penelitian yang sudah terindeks atau terakreditasi nasional *Science and Technology Index* (SINTA). SINTA merupakan portal *online* yang berkaitan dengan pengukuran kinerja suatu Ilmu Pengetahuan dan Teknologi (IPTEK) tentang pengukuran kinerja dosen atau peneliti, penulis, penulis buku, kinerja jurnal dan kinerja institute IPTEK tersebut. Pada SINTA terdapat beberapa tingkatan atau klasifikasi jurnal yang terbagi menjadi enam kategori yaitu SINTA 1,2,3,4,5 dan 6, dimana tingkatan SINTA 1 merupakan tingkatan yang tertinggi dan SINTA 6 yang terendah. SINTA hanya berisikan tentang database jurnal yang ada di Indonesia yang dikelompok berdasarkan kategori keilmuan. SINTA tidak berfungsi dalam mengidentifikasi suatu topik penelitian tapi hanya fokus pada identifikasi jurnal penelitian yang ada di Indonesia. Sedangkan pembaca perlu suatu data yang dapat dijadikan rujukan yang dapat digunakan dalam penentuan topik penelitian, salah satunya pada topik penelitian Pendidikan Matematika.

Pada topik penelitian Pendidikan Matematika pembaca atau peneliti masih mengalami kesulitan dalam mengidentifikasi topik penelitian di bidang Pendidikan Matematika. Hal ini disebabkan karena belum adanya sistem atau model yang dapat dilihat atau digunakan dalam penentuan topik penelitian. Disamping itu belum adanya otomatisasi arah penelitian Pendidikan Matematika di Indonesia dengan menggunakan pemodelan topik, sehingga perlu dilakukan kajian atau penelitian tentang ini.

Berdasarkan kajian literature belum adanya sistem yang dibangun untuk merekomendasikan topik penelitian di bidang Pendidikan Matematika. dari beberapa sumber yang dibaca belum adanya sistem yang dibangun atau digunakan untuk mempermudah peneliti dalam menentukan topik penelitian yang akan dilakukan sebagai langkah awal dalam memulai penelitian. Dalam penelitian suatu hal yang terpenting

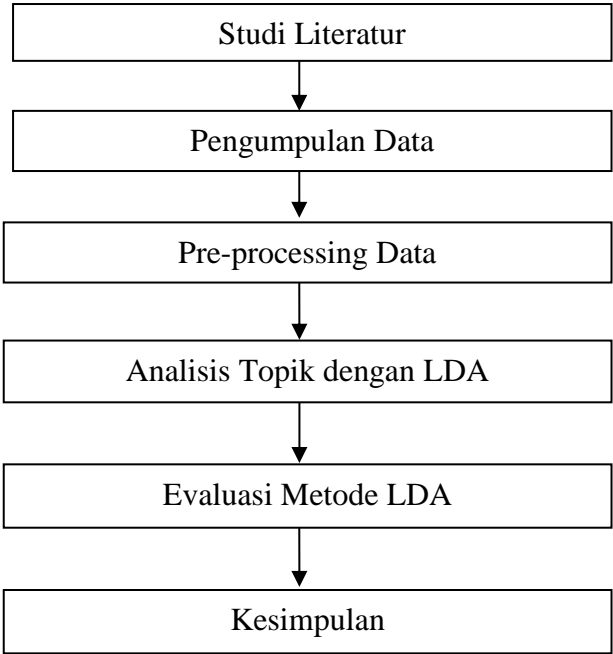
adalah trend penelitian yang sedang berkembang sehingga dapat menentukan kebaharuan (*novelty*) dari penelitian-penelitian yang sudah dilakukan sebelumnya. Sedangkan belum adanya sistem yang digunakan untuk menentukan trend dan *state of the art* dari penelitian di bidang Pendidikan Matematika. Untuk itu perlu dilakukan penelitian tentang analisis topik penelitian Pendidikan Matematika yang dikenal dengan pemodelan topik.

Pemodelan topik merupakan suatu pendekatan untuk menganalisis kumpulan dokumen berbentuk teks dan mengelompokkan menjadi beberapa topik. Pemodelan topik dapat dilakukan dengan beberapa metode, salah satunya metode *Latent Dirichlet Allocation* (LDA). LDA merupakan metode untuk pemodelan topik yang berdasarkan pada konsep probabilitas yang bertujuan untuk mencari kemiripan suatu dokumen dan mengelompokkan dokumen-dokumen menjadi beberapa topik atau kelompok [16]. Dengan adanya pemodelan topik dengan metode LDA dapat mengatasi permasalahan tentang bagaimana menentukan topik penelitian Pendidikan Matematika di Indonesia.

2. METODE[d1]

Tahapan awal yang dilakukan dalam penelitian ini adalah dengan studi literature dimana dilakukan pengumpulan dan analisis sumber literatur yang terkait tentang metode *Latent Dirichlet Allocation* (LDA) dan penelitian pendidikan matematika yang meliputi buku referensi, jurnal, sumber data primer maupun sekunder, dan artikel prosiding. Tahapan pengumpulan data dilakukan dengan pengambilan data pada jurnal nasional Pendidikan Matematika yang sudah terakreditasi SINTA (*Science and Technology Index*) yaitu SINTA 1,2,3 dan 4 dimana merupakan artikel-artikel penelitian yang berkaitan dengan penelitian pendidikan matematika. Tahapan berikutnya yakni *pre-processing* data dimana dilakukan proses cleaning data untuk terhindar dari bias [17]. Setelah melewati *preprocessing* data, selanjutnya data dianalisis menggunakan model LDA. Hasil dari model yang dibuat kemudian di evaluasi dengan melakukan perhitungan nilai *perplexity* serta *coherence* [18]. Evaluasi dilakukan untuk mengetahui keakuratan dari validasi topik [19]. *Perplexity* adalah ukuran kinerja dari pemodelan bahasa berdasarkan probabilitas rata-rata yang telah dikembangkan [20]. Pada analisis *Topic Modeling* dengan *Latent Dirichlet Allocation*, peneliti dalam melakukan analisis, packages yang digunakan dari python diantaranya yaitu packages Gensim dan pyLDAvis.

Perancangan penelitian dilakukan dengan tahapan sebagai berikut. [d2]



Gambar 1. Rancangan Penelitian

3. **HASIL DAN PEMBAHASAN**[d3]

3.1. **Hasil**[d4]

Hasil penelitian dijabarkan sebagai berikut.

3.1.1. *Pre-processing*

Pada tahap preprocessing dilakukan untuk menghilangkan kata-kata dalam dokumen yang tidak dibutuhkan. Tahap ini merupakan tahap pertama yang dilakukan dalam sebelum mulai melakukan analisis. Berikut merupakan beberapa data yang digunakan.

Tabel 1. **Data Awal**

Abstrak
Rendahnya kemampuan pemahaman konsep siswa pada materi operasi hitung bentuk aljabar menjadi salah satu permasalahan yang ditemui saat belajar materi Aljabar. Banyak miskonsepsi yang ditemukan pada materi operasi hitung bentuk aljabar. Tujuan dari penelitian ini adalah untuk mengetahui bentuk miskonsepsi yang dialami siswa pada materi operasi hitung bentuk aljabar.

Pada tahap *preprocessing* ini terdiri dari beberapa proses seperti *case folding*, *remove punctuation*, *stopword*, dan *tokenizing*. Proses *preprocessing* dikerjakan menggunakan *Tools Rapid Miner Studio*. mengerjakan proses *preprocessing* terdapat 4 tahapan proses untuk memperoleh hasil yang maksimal, sebagai berikut.

Berikut proses *cleaning* kode program yang mengimplementasikan *cleaning* data dapat dilihat pada gambar di bawah ini.

```

import re
import string
import nltk
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords

nltk.download('stopwords')
nltk.download('punkt')

factory = StemmerFactory()
stemmer = factory.create_stemmer()

stop_words_indo = set(stopwords.words('indonesian'))
stop_words_eng = set(stopwords.words('english'))
stop_words = stop_words_indo.union(stop_words_eng)

def preprocessing(text):
    tokens = word_tokenize(text)
    tokens = [word.lower() for word in tokens]
    tokens = [re.sub(r'[^a-zA-Z]', '', word) for word in tokens]
    tokens = [word for word in tokens if not word in stop_words]
    tokens = [stemmer.stem(word) for word in tokens]

    # Tokenisasi kembali (Opsional)
    text = ' '.join(tokens)

    return text

```

Gambar 2. Proses *Cleaning*

```

Abstrak \
0 Rendahnya kemampuan pemahaman konsep siswa pad...
1 Sebagian besar siswa masih mengalami kesulitan...
2 Berbagai penelitian mengemukakan kemampuan ber...
3 Penelitian ini bermaksud untuk memperoleh hasi...
4 Rendahnya kemampuan penalaran menjadi penyebab...
...
1010 Penelitian ini bertujuan untuk menghasilkan mo...
1011 Kontinuitas dan diferensiabilitas merupakan ko...
1012 Penelitian ini bertujuan untuk mendeskripsikan...
1013 Penelitian ini bertujuan untuk mengetahui besa...
1014 Penelitian ini bertujuan untuk mendeskripsikan...

Preprocessed_Abstrak
0 rendah mampu paham konsep siswa materi operasi...
1 siswa alami sulit ajar kerucut desain lintas ...
2 teliti muka mampu pikir kreatif matematis sisw...
3 teliti oleh hasil tingkat mampu koneksi matema...
4 rendah mampu nalar sebab rendah prestasi matem...
...
1010 teliti tuju hasil modul ajar bas materi trigon...
1011 kontinuitas diferensiabilitas konsep asai maha...
1012 teliti tuju deskripsi terampil metakognitif si...
1013 teliti tuju pengaruh beda pengaruh guna media ...
1014 teliti tuju deskripsi efektif model ajar sbl m...

```

Gambar 3. Hasil *Cleaning*

Pada gambar di atas adalah tampilan Hasil *Cleaning*, dimana data sudah bersih untuk proses selanjutnya.

Case folding merupakan proses mengubah karakter dari huruf besar menjadi huruf kecil dengan tujuan jika terdapat kata yang sama namun penulisannya berbeda huruf kapital atau tidak maka tidak dianggap kata yang berbeda. Hasil dari *case folding* yaitu sebagai berikut.

Tabel 2. Hasil *Case Folding*

Abstrak

rendahnya kemampuan pemahaman konsep siswa pada materi operasi hitung bentuk aljabar menjadi salah satu permasalahan yang ditemui saat belajar materi aljabar. banyak miskonsepsi yang ditemukan pada materi operasi hitung bentuk aljabar. tujuan dari penelitian ini adalah untuk mengetahui bentuk miskonsepsi yang dialami siswa pada materi operasi hitung bentuk aljabar.

Remove punctuation merupakan proses yang dilakukan untuk membuang karakter seperti tanda baca yang tidak digunakan seperti tabel berikut.

Tabel 3. Hasil dari Remove Punctuation

Sebelum	Sesudah
rendahnya kemampuan pemahaman konsep siswa pada materi operasi hitung bentuk aljabar menjadi salah satu permasalahan yang ditemui saat belajar materi aljabar. banyak miskonsepsi yang ditemukan pada materi operasi hitung bentuk aljabar. tujuan dari penelitian ini adalah untuk mengetahui bentuk miskonsepsi yang dialami siswa pada materi operasi hitung bentuk aljabar.	rendahnya kemampuan pemahaman konsep siswa pada materi operasi hitung bentuk aljabar menjadi salah satu permasalahan yang ditemui saat belajar materi aljabar banyak miskonsepsi yang ditemukan pada materi operasi hitung bentuk aljabar tujuan dari penelitian ini adalah untuk mengetahui bentuk miskonsepsi yang dialami siswa pada materi operasi hitung bentuk aljabar

Stopwords merupakan kata-kata yang sering muncul dalam dokumen teks dan kata dalam dokumen tersebut tidak berguna. Setiap kata yang muncul di dalam corpus sering mempunyai banyak varian morfologik. Oleh karena itu, setiap kata-kata direduksi ke bentuk *stemmed word (term)* yang cocok. Kata-kata tersebut akan diambil bentuk kata dasarnya dengan cara menghilangkan awalan atau akhiran seperti tabel berikut.

Tabel 4. Hasil dari Stopwords & Stemming

Sebelum	Sesudah
rendahnya kemampuan pemahaman konsep siswa pada materi operasi hitung bentuk aljabar menjadi salah satu permasalahan yang ditemui saat belajar materi aljabar banyak miskonsepsi yang ditemukan pada materi operasi hitung bentuk aljabar tujuan dari penelitian ini adalah untuk mengetahui bentuk miskonsepsi yang dialami siswa pada materi operasi hitung bentuk aljabar	rendah mampu paham konsep siswa materi operasi hitung bentuk aljabar salah masalah temu ajar materi aljabar miskonsepsi temu materi operasi hitung bentuk aljabar tuju teliti bentuk miskonsepsi alami siswa materi operasi hitung bentuk aljabar

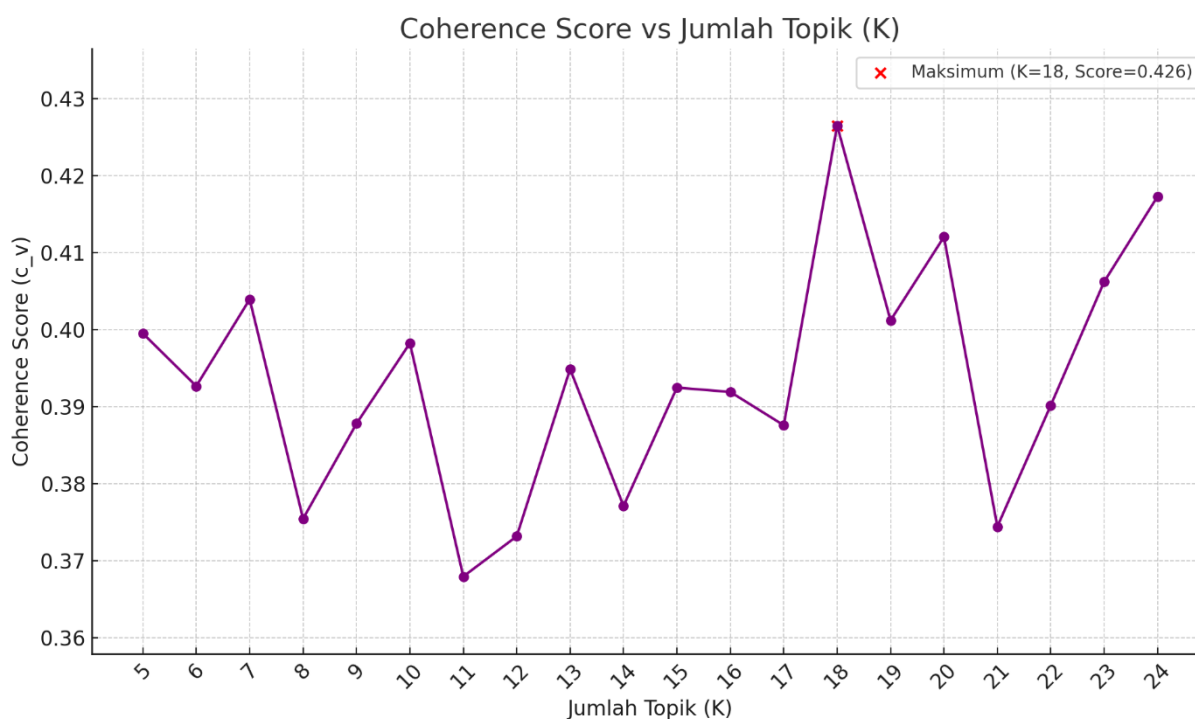
Tahapan *tokenizing* merupakan tahap yang berfungsi untuk memisahkan antara kata yang terdapat pada corups. Proses *tokenizing* menggunakan spasi sebagai karakter untuk tanda pemisah setiap katanya seperti tabel berikut.

Tabel 5. Hasil dari *tokenizing*

Sebelum	Sesudah
rendah mampu paham konsep siswa materi operasi hitung bentuk aljabar salah masalah temu ajar materi aljabar miskonsepsi temu materi operasi hitung bentuk aljabar tuju teliti bentuk miskonsepsi alami siswa materi operasi hitung bentuk aljabar	['rendah', 'mampu', 'paham', 'konsep', 'siswa', 'materi', 'operasi', 'hitung', 'bentuk', 'aljabar', 'salah', 'masalah', 'temu', 'ajar', 'materi', 'aljabar', 'miskonsepsi', 'temu', 'materi', 'operasi', 'hitung', 'bentuk', 'aljabar', 'uju', 'teliti', 'bentuk', 'miskonsepsi', 'alami', 'siswa', 'materi', 'operasi', 'hitung', 'bentuk', 'aljabar']

3.1.2. Evaluasi Topic Modeling dengan *Latent Dirichlet Allocation* (LDA)

Dalam menentukan hasil pemodelan dapat dilakukan dengan melihat pada visualisasi dari grafik *coherence score*. *Coherence score* merupakan suatu ukuran yang akan digunakan untuk mengevaluasi *Topic Modeling*, dimana jika *coherence* skor topik yang tinggi maka model yang dihasilkan tersebut yang baik. Berikut hasil *Coherence score* dari topik penelitian pendidikan matematika yang diambil dari abstrak publikasi ilmiah.

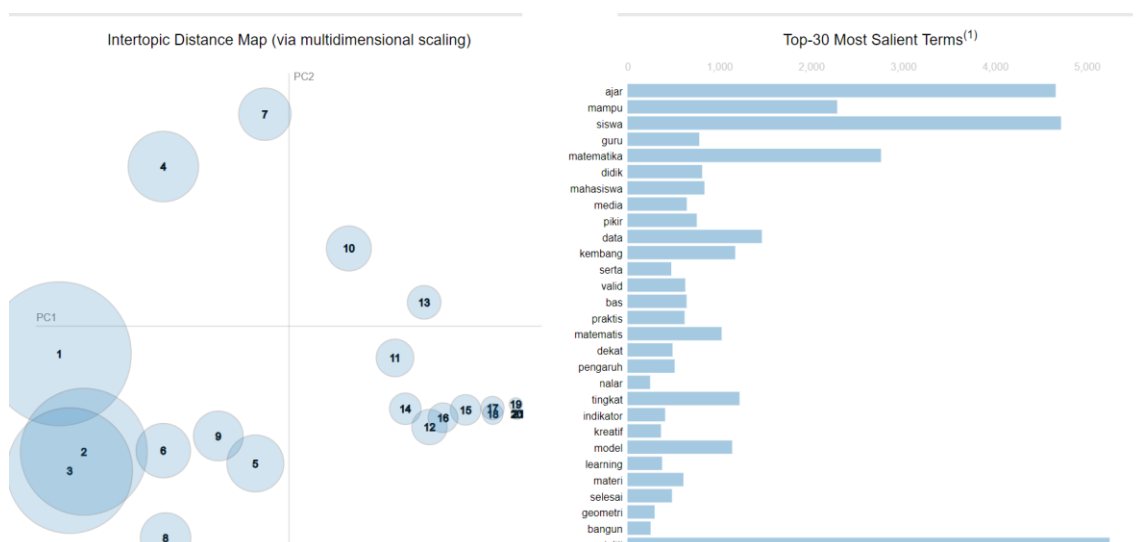


Grafik 1. *Coherence score*

Berdasarkan grafik *coherence score* diperoleh jumlah topik terbaik (K) = 18, dengan *coherence score* = 0.426 (tertinggi). Ini menunjukkan bahwa model LDA menghasilkan topik yang paling koheren secara semantik saat menggunakan 18 topik.

3.1.3. Hasil Topic Modeling dengan Latent Dirichlet Allocation (LDA)

Berikut hasil *topic modeling* LDA.



Gambar 4. Hasil *topic modeling*

Topik 1:

(1, '0.367*"didik" + 0.296*"serta" + 0.036*"linier" + 0.034*"periksa" + '0.031*"program" + 0.026*"fungsi" + 0.011*"efficacy" + 0.009*"self" + '0.008*"wajib" + 0.004*"explaining")',

Interpretasi Model LDA Topik 1: *Self Efficacy*

Topik 2:

(2, '0.122*"kembang" + 0.098*"valid" + 0.097*"praktis" + 0.043*"kriteria" + '0.043*"hasil" + 0.036*"bahan" + 0.034*"efektif" + 0.032*"validasi" + '0.029*"angket" + 0.028*"rata-rata")',

Interpretasi Model LDA Topik 2: Validitas, Praktikalitas, efektivitas bahan ajar

Topik 3:

(3, '0.200*"mahasiswa" + 0.180*"pikir" + 0.087*"kreatif" + 0.070*"kritis" + '0.035*"tingkat" + 0.025*"didik" + 0.024*"kuliah" + 0.024*"kreativitas" + '0.019*"program" + 0.019*"ikut")'),

Interpretasi Model LDA Topik 3: Berfikir kritis dan kreatif

Topik 4:

(4, '0.097*"indikator" + 0.092*"data" + 0.072*"gaya" + 0.043*"siswa" + '0.038*"selesai" + 0.037*"prosedur" + 0.034*"penuh" + 0.033*"simpul" + '0.031*"wawancara" + 0.030*"kognitif")'),

Interpretasi Model LDA Topik 4: Gaya kognitif

Topik 5:

(5, '0.078*"konteks" + 0.063*"perangkat" + 0.058*"learning" + 0.041*"desain" + '0.040*"aktivitas" + 0.032*"research" + 0.031*"luas" + 0.028*"hidup" + '0.026*"based" + 0.023*"problem")'),

Interpretasi Model LDA Topik 5: *Problem based learning*

Topik 6:

(6, '0.000*"ordinal" + 0.000*"guided inquiry" + 0.000*"employing" + '0.000*"sempls" + 0.000*"lelah" + 0.000*"bootstrap" + 0.000*"shops" + '0.000*"prof" + 0.000*"functional" + 0.000*"hamka")'),

Interpretasi Model LDA Topik 6: *Guided Inquiry*

Topik 7:

(7, '0.093*"geometri" + 0.054*"data" + 0.050*"tahap" + 0.044*"teliti" + '0.039*"level" + 0.036*"pisa" + 0.036*"transformasi" + 0.024*"liput" + '0.023*"teknik" + 0.020*"sekolah")'),

Interpretasi Model LDA Topik 7: Level geometri PISA

Topik 8:

(8, '0.142*"problem" + 0.123*"solving" + 0.050*"creative" + 0.033*"mind" + '0.031*"habits" + 0.000*"ctl" + 0.000*"efektifitas" + 0.000*"diskusi" + '0.000*"videos" + 0.000*"asistif")'),

Interpretasi Model LDA Topik 8: *Problem solving dan Habits of mind*

Topik 9:

(9, '0.139*"ukur" + 0.101*"kelompok" + 0.060*"dekat" + 0.041*"yogyakarta" + '0.036*"tinjau" + 0.034*"studi" + 0.030*"vii" + 0.026*"buat" + '0.025*"percaya" + 0.022*"saintifik")'),

Interpretasi Model LDA Topik 9: Studi *Saintifik*

Topik 10:

(10, '0.077*"matematika" + 0.067*"teliti" + 0.027*"hasil" + 0.025*"paham" + '0.024*"siswa" + 0.022*"konsep" + 0.020*"dasar" + 0.020*"proses" + '0.019*"salah" + 0.016*"temu")'),

Interpretasi Model LDA Topik 10: Pemahaman konsep matematis siswa

Topik 11:

(11, '0.404*"media" + 0.052*"mathematics" + 0.052*"interaktif" + 0.049*"online" + "0.041*"stem" + 0.037*"digital" + 0.029*"akibat" + 0.025*"research" + "0.019*"didik" + 0.018*"tarik")'),

Interpretasi Model LDA Topik 11: Media Interaktif matematika

Topik 12:

(12, '0.093*"bas" + 0.067*"ahli" + 0.064*"modul" + 0.053*"materi" + '0.049*"respon" + 0.047*"coba" + 0.046*"kembang" + 0.044*"layak" + '0.037*"development" + 0.037*"nyata")'),

Interpretasi Model LDA Topik 12: *Development* modul

Topik 13:

(13, '0.142*"cerdas" + 0.083*"logis" + 0.045*"prediksi" + 0.021*"mmp" + '0.000*"emosional" + 0.000*"spasial" + 0.000*"kpm" + 0.000*"sekuensial" + '0.000*"sc" + 0.000*"derajat")'),

Interpretasi Model LDA Topik 13: Kecerdasan logis dan spasial

Topik 14:

(14, '0.140*"siswa" + 0.109*"mampu" + 0.069*"teliti" + 0.050*"matematis" + "0.035*"pecah" + 0.024*"tes" + 0.019*"rendah" + 0.019*"data" + '0.018*"terampil" + 0.018*"smp")'),

Interpretasi Model LDA Topik 14: Kemampuan matematis siswa

Topik 15:

(15, '0.457*"nalar" + 0.064*"regresi" + 0.046*"ganda" + 0.045*"selatan" + '0.015*"sejarah" + 0.012*"sumatera" + 0.007*"museum" + '0.005*"selfconfidence" + 0.005*"kota" + 0.005*"pustaka")'),

Interpretasi Model LDA Topik 15: *Self Confidence*

Topik 16:

(16, '0.078*"kompetensi" + 0.077*"rpp" + 0.074*"buku" + 0.067*"integrasi" + "0.055*"tugas" + 0.047*"sosial" + 0.041*"kurikulum" + 0.036*"teks" + '0.034*"fase" + 0.024*"plomp")'),

Interpretasi Model LDA Topik 16: Kurikulum Matematika

Topik 17:

(17, '0.162*"ajar" + 0.074*"matematika" + 0.071*"pengaruh" + 0.058*"literasi" + "0.057*"mandiri" + 0.047*"daring" + 0.042*"signifikan" + 0.034*"motivasi" + "0.024*"pbl" + 0.020*"lingkung")'),

Interpretasi Model LDA Topik 17: Literasi matematika

Topik 18:

(18, '0.155*"budaya" + 0.143*"minat" + 0.094*"etnomatematika" + 0.064*"kerangka" + 0.041*"kerja" + 0.029*"cermin" + 0.024*"geser" + 0.021*"ilmu" + '0.018*"mts" + 0.016*"aritmetika")'),

Interpretasi Model LDA Topik 18: Etnomatematika

3.2. Pembahasan[d5]

Pemodelan topik dengan *latent dirichlet allocation* (LDA) menggunakan data abstrak publikasi ilmiah Pendidikan Matematika di Indonesia sejumlah 1015 artikel yang diperoleh dari SINTA 2 dari tahun 2020-2023. Tahap awal yang dilakukan adalah tahap *pre-processing* yang terdiri dari *case folding*, pada tahap *case folding* dilakukan merupakan proses mengubah karakter dari huruf besar menjadi huruf kecil dengan tujuan jika terdapat kata yang sama namun penulisannya berbeda huruf kapital atau tidak maka tidak dianggap kata yang berbeda. Pada tahap ini, semua huruf akan diubah menjadi *lowercase* atau huruf kecil. Setelah proses ini dilakukan dilanjutkan dengan tahap *remove punctuation*, pada tahap *remove punctuation* merupakan proses yang dilakukan untuk membuang karakter seperti tanda baca yang tidak digunakan.

Langkah selanjutnya adalah tahap *stopwords removal (filtering) & stemming*, pada tahap ini dilakukan pemeriksaan terhadap kata-kata yang sering muncul dalam dokumen teks dan kata dalam dokumen tersebut yang tidak berguna. Apabila dokumen tersebut mengandung kata sambung, kata depan, kata ganti atau kata yang tidak ada hubungannya dengan pemodelan topik, maka kata tersebut akan dihapus, berikutnya dilakukan *stemming* dengan cara kata-kata tersebut akan diambil bentuk kata dasarnya dengan cara menghilangkan awalan atau akhiran, tahap *tokenizing*, tahapan *tokenizing* merupakan tahap yang berfungsi untuk memisahkan antara kata yang terdapat pada corpus dan dilanjutkan dengan proses pemodelan topik menggunakan LDA dengan menentukan jumlah topik berdasarkan *coherence score*.

Hasil dari analisis pemodelan topik dengan metode *latent dirichlet allocation* (LDA) pada data abstrak artikel ilmiah Pendidikan Matematika pada Jurnal terakreditasi SINTA 2 di Indonesia dari tahun 2020-2023 diperoleh jumlah topik sebanyak 18 dengan *coherence score* sebesar 0.426 dengan kata yang saling berkaitan. *Coherence score* merupakan suatu ukuran yang akan digunakan untuk mengevaluasi *topic modeling*, dimana jika *coherence* skor topik yang tinggi maka model yang dihasilkan tersebut yang baik.

Berdasarkan evaluasi topik pemodelan dengan LDA dihasilkan [d6]beberapa topik yaitu: berdasarkan model LDA topik 1 didapatkan hasil kata-kata yang sering muncul dan terkait satu sama lain yaitu kata-kata didik, serta, linier, periksa, program, fungsi, efficacy, self, wajib, dan explaining. Berdasarkan beberapa data yang diperoleh dapat

disimpulkan bahwa pemodelan LDA topik 1 membahas tentang *Self efficacy* pada materi program linier dan fungsi yang berkaitan dengan memeriksa dan *explaining* terhadap solusi dari penyelesaian soal yang diberikan.

Berdasarkan model LDA topik 2 didapatkan hasil kata-kata yang sering muncul dan terkait satu sama lain yaitu kembang, valid, praktis, kriteria, hasil, bahan, efektif, validasi, angket, dan rata-rata. Berdasarkan beberapa data pemodelan LDA topik 2 yang diperoleh dapat disimpulkan bahwa pengembangan bahan ajar dilihat dari rata-rata hasil validitas, praktikalitas, dan efektivitas yang diperoleh dari angket yang diberikan kepada siswa. Berdasarkan model LDA topik 3 didapatkan hasil kata-kata yang sering muncul dan terkait satu sama lain yaitu mahasiswa, pikir, kreatif, kritis, tingkat, didik, kuliah, kreativitas, program, dan ikut. Berdasarkan beberapa data pemodelan LDA topik 3 yang diperoleh dapat disimpulkan bahwa proses berfikir kritis dan kreatif berhubungan dengan proses berfikir yang mengembangkan kreativitas tingkat tinggi mahasiswa dan merancang program penyelesaian masalah terhadap peserta didik.

Berdasarkan model LDA topik 4 didapatkan hasil kata-kata yang sering muncul dan terkait satu sama lain yaitu Indikator, data, gaya, siswa, selesai, prosedur, penuh, simpul, wawancara, dan kognitif. Berdasarkan beberapa data pemodelan LDA topik 4 yang diperoleh dapat disimpulkan bahwa data gaya kognitif siswa dapat dilihat dari indikator dan prosedur pengumpulan data gaya kognitif berupa wawancara sehingga mencerminkan hasil yang mendalam dan diperoleh data sampai selesai. Berdasarkan model LDA topik 5 didapatkan hasil kata-kata yang sering muncul dan terkait satu sama lain yaitu konteks, perangkat, learning, desain, aktivitas, research, luas, hidup, based, dan problem. Berdasarkan beberapa data pemodelan LDA topik 5 yang diperoleh dapat disimpulkan bahwa *problem based learning* merupakan model pembelajaran yang dapat diterapkan dalam research desain perangkat pembelajaran dan pada aktivitas materi luas.

Berdasarkan model LDA topik 6 didapatkan hasil kata-kata yang sering muncul dan terkait satu sama lain yaitu ordinal, guided inquiry, employing, sempls, lelah, bootstrap, shops, prof, functional, dan hamka. Berdasarkan beberapa data pemodelan LDA topik 6 yang diperoleh dapat disimpulkan bahwa topik penelitian *guided inquiry* dapat dilakukan di perguruan tinggi maupun sekolah termasuk pada materi functional ordinal. Berdasarkan model LDA topik 7 didapatkan hasil kata-kata yang sering muncul dan terkait satu sama lain yaitu geometri, data, tahap, teliti, level, pisa, transformasi, liput,

teknik, dan sekolah. Berdasarkan beberapa data pemodelan LDA topik 7 yang diperoleh dapat disimpulkan bahwa geometri transformasi diajarkan pada mahasiswa teknik dan sekolah yang disesuaikan dengan tahapan level geometri PISA.

Berdasarkan model LDA topik 8 didapatkan hasil kata-kata yang sering muncul dan terkait satu sama lain yaitu problem, solving, creative, mind, habits, dll, efektifitas, diskusi, videos, dan asistif. Berdasarkan beberapa data pemodelan LDA topik 8 yang diperoleh dapat disimpulkan bahwa *Problem solving* dan *Habits of mind* merupakan topik pembelajaran yang dapat meningkatkan efektivitas dan creative melalui pengamatan videos dan diskusi. Berdasarkan model LDA topik 9 didapatkan hasil kata-kata yang sering muncul dan terkait satu sama lain yaitu ukur, kelompok, dekat, Yogyakarta, tinjau, studi, vii, buat, percaya, dan saintifik. Berdasarkan beberapa data pemodelan LDA topik 9 yang diperoleh dapat disimpulkan bahwa studi *saintifik* pada materi kelas VII menggunakan konteks budaya Yogyakarta.

Berdasarkan model LDA topik 10 didapatkan hasil kata-kata yang sering muncul dan terkait satu sama lain yaitu matematika, teliti, hasil, paham, siswa, konsep, dasar, proses, salah, dan temu. Berdasarkan beberapa data pemodelan LDA topik 10 yang diperoleh dapat disimpulkan bahwa pemahaman konsep matematis siswa merupakan topik penelitian yang menitikberatkan pada proses pemahaman konsep dasar dan menemukan kesalahan dalam memahami konsep. Berdasarkan model LDA topik 11 didapatkan hasil kata-kata yang sering muncul dan terkait satu sama lain yaitu media, mathematics, interaktif, online, stem, digital, akibat, research, didik dan tarik. Berdasarkan beberapa data pemodelan LDA topik 11 yang diperoleh dapat disimpulkan bahwa media interaktif mathematics merupakan media pembelajaran digital yang dapat dilakukan secara online dan berbasis STEM merupakan topik penelitian yang dapat meningkatkan aktivitas dan daya tarik peserta didik untuk belajar.

Berdasarkan model LDA topik 12 didapatkan hasil kata-kata yang sering muncul dan terkait satu sama lain yaitu bas, ahli, modul, materi, respon, coba, kembang, layak, development dan nyata. Berdasarkan beberapa data pemodelan LDA topik 12 yang diperoleh dapat disimpulkan bahwa topik *development* modul dilakukan dengan cara uji coba dan validasi ke validator untuk melihat kelayakan modul. Berdasarkan model LDA topik 13 didapatkan hasil kata-kata yang sering muncul dan terkait satu sama lain yaitu cerdas, logis, prediksi, mmp, emosional, spasial, kpm, sekuensial, sc, dan derajat.

Berdasarkan beberapa data pemodelan LDA topik 13 yang diperoleh dapat disimpulkan bahwa topik kecerdasan logis dan spasial yang dapat dihubungkan dengan kecerdasan emosional.

Berdasarkan model LDA topik 14 didapatkan hasil kata-kata yang sering muncul dan terkait satu sama lain yaitu siswa, mampu, teliti, matematis, pecah, tes, rendah, data, terampil dan smp. Berdasarkan beberapa data pemodelan LDA topik 14 yang diperoleh dapat disimpulkan bahwa topik kemampuan matematis siswa pada materi pecahan dapat meningkatkan keterampilan dan ketelitian dalam memecahkan masalah matematis. Berdasarkan model LDA topik 15 didapatkan hasil kata-kata yang sering muncul dan terkait satu sama lain yaitu nalar, regresi, ganda, selatan, sejarah, sumatera, museum, selfconfidence, kota, dan pustaka. Berdasarkan beberapa data pemodelan LDA topik 15 yang diperoleh dapat disimpulkan bahwa topik penalaran dan *self confidence* merupakan topik penelitian pada pembelajaran matematika yang sering dikaitkan dengan konteks budaya dengan cara analisis regresi linear berganda.

Berdasarkan model LDA topik 16 didapatkan hasil kata-kata yang sering muncul dan terkait satu sama lain yaitu kompetensi, rpp, buku, integrasi, tugas, sosial, kurikulum, teks, fase dan plomp. Berdasarkan beberapa data pemodelan LDA topik 16 yang diperoleh dapat disimpulkan bahwa topik pengembangan perangkat pembelajaran merupakan bagian dari kajian kurikulum matematika yang terdiri pengembangan kompetensi matematis, RPP, buku dan integrasi tugas siswa, prosedur pengembangan yang sering digunakan berupa fase Plomp. Berdasarkan model LDA topik 17 didapatkan hasil kata-kata yang sering muncul dan terkait satu sama lain yaitu ajar, matematika, pengaruh, literasi, mandiri, daring, signifikan, motivasi, pbl, dan lingkung. Berdasarkan beberapa data pemodelan LDA topik 17 yang diperoleh dapat disimpulkan bahwa topik literasi matematika, motivasi, dan kemandirian belajar yang dilakukan dengan cara menentukan pengaruhnya secara signifikan antar komponen. Berdasarkan model LDA topik 18 didapatkan hasil kata-kata yang sering muncul dan terkait satu sama lain yaitu budaya, minat, etnomatematika, kerangka, kerja, cermin, geser, ilmu, mts, dan aritmetika. Berdasarkan beberapa data pemodelan LDA topik 18 yang diperoleh dapat disimpulkan bahwa topik etnomatematika dapat dihubungkan dengan minat belajar siswa di MTs pada materi aritmatika sosial.

4. KESIMPULAN[d7]

Pemodelan topik dengan *latent dirichlet allocation* (LDA) menggunakan data abstrak publikasi ilmiah Pendidikan Matematika di Indonesia sejumlah 1015 artikel yang diperoleh dari SINTA 2 dari tahun 2020-2023. Selanjutnya dilakukan tahap *pre-processing* yang terdiri dari *case folding*, *remove punctuation*, *stopword*, dan *tokenizing* dan dilanjutkan dengan proses pemodelan topik menggunakan LDA dengan menentukan jumlah topik berdasarkan *coherence score*.

Hasil dari analisis pemodelan topik dengan metode *latent dirichlet allocation* (LDA) pada data abstrak artikel ilmiah Pendidikan Matematika pada Jurnal terakreditasi SINTA 2 di Indonesia dari tahun 2020-2023 diperoleh jumlah topik sebanyak 18 dengan *coherence score* sebesar 0.426 dengan kata yang saling berkaitan.

Berdasarkan evaluasi topik pemodelan dengan LDA dihasilkan kesimpulan yaitu: Model LDA Topik ke-1: *Self Efficacy*, Model LDA Topik ke-2: Validitas, Praktikalitas, efektivitas bahan ajar, Model LDA Topik ke-3: Berfikir kritis dan kreatif, Model LDA Topik ke-4: Gaya kognitif, Model LDA Topik ke-5: *Problem based learning*, Model LDA Topik ke-6: *Guided Inquiry*, Model LDA Topik ke-7: Level geometri PISA, Model LDA Topik ke-8: *Problem solving dan Habits of mind*, Model LDA Topik ke-9: Studi Saintifik, Model LDA Topik ke-10: Pemahaman konsep matematis siswa, Model LDA Topik ke-11: Media Interaktif matematika, Model LDA Topik ke-12: *Development* modul, Model LDA Topik ke-13: Kecerdasan logis dan spasial, Model LDA Topik ke-14: Kemampuan matematis siswa, Model LDA Topik ke-15: *Self Confidence*, Model LDA Topik ke-16: Kurikulum Matematika, Model LDA Topik ke-17: Literasi matematika, Model LDA Topik ke-18: Etnomatematika.

5. DAFTAR PUSTAKA

- [1] L. Hickman, S. Thapa, L. Tay, M. Cao, and P. Srinivasan, "Text Preprocessing for Text Mining in Organizational Research: Review and Recommendations," *Organ. Res. Methods*, vol. 25, no. 1, 2022, doi: 10.1177/1094428120971683.
- [2] S. Kumar, A. K. Kar, and P. V. Ilavarasan, "Applications of text mining in services management: A systematic literature review," *International Journal of Information Management Data Insights*, vol. 1, no. 1, 2021, doi: 10.1016/j.jjime.2021.100008.
- [3] M. Alzate, M. Arce-Urriaza, and J. Cebollada, "Mining the text of online consumer reviews to analyze brand image and brand positioning," *J. Retail. Consum. Serv.*, vol. 67, 2022, doi: 10.1016/j.jretconser.2022.102989.
- [4] F. S. Ramadhan and A. P. W. Wibowo, "Text Clustering Analysis Of Public Sentiment

- On Twitter About Tokopedia Bts Using Orange Data Mining,” *J. Darma Agung*, Vol. Jurnal Dar, 2022.
- [5] A. Deolika, K. Kusriani, and E. T. Luthfi, “Analisis Pembobotan Kata Pada Klasifikasi Text Mining,” *J. Teknol. Inf.*, vol. 3, no. 2, 2019, doi: 10.36294/jurti.v3i2.1077.
- [6] A. Y. Muniar, P. Pasnur, and K. R. Lestari, “Penerapan Algoritma K-Nearest Neighbor pada Pengklasifikasian Dokumen Berita Online,” *Inspir. J. Teknol. Inf. dan Komun.*, vol. 10, no. 2, 2020, doi: 10.35585/inspir.v10i2.2570.
- [7] R. Gusdiana, I. Alfian, and C. Juliane, “Implementation Of Text Processing For Sentiment Analysis Of Tax Payment Interest After The ‘Rubicon’ Phenomenon,” *J. Tek. Inform.*, vol. 4, no. 5, 2023, doi: 10.52436/1.jutif.2023.4.5.1014.
- [8] K. Rosikin, S. Basuki, and Y. Azhar, “Ekstraksi Informasi Kesehatan Masyarakat Dari Tweet Berbahasa Indonesia Berbasis Klasifikasi Dengan Algoritma Naive Bayes,” *J. Repos.*, vol. 2, no. 2, 2020, doi: 10.22219/repositor.v2i2.237.
- [9] F. Yao and Y. Wang, “Tracking urban geo-topics based on dynamic topic model,” *Comput. Environ. Urban Syst.*, vol. 79, 2020, doi: 10.1016/j.compenvurbsys.2019.101419.
- [10] Y. F. Hernawan, P. P. Adikara, and R. C. Wihandika, “Peringkasan Artikel Berbahasa Indonesia Menggunakan TextRank dengan Pembobotan BM25,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 1, 2022, doi: 10.25126/jtiik.2022913765.
- [11] M. Noperia, I. Ishak, and V. W. Sari, “Implementasi Data Mining Pengelompokan Data Nilai Untuk Menentukan Minat Belajar Seni Budaya,” *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, vol. 2, no. 1, 2023, doi: 10.53513/jursi.v2i1.5656.
- [12] D. Nurmalasari and H. Ribut Yuliantoro, “Implementasi Ekstraksi Fitur untuk Pengelompokan Dokumen Proposal Menggunakan Algoritma Naïve Bayes,” *J. Komput. Terap.*, vol. 8, no. 1, 2022, doi: 10.35143/jkt.v8i1.5351.
- [13] M. A. P. Subali and P. Wijaya, “Sistem Question Answering untuk Bahasa Bali menggunakan Metode Rule-Based dan String Similarity,” *Techno.Com*, vol. 20, no. 2, 2021, doi: 10.33633/tc.v20i2.4390.
- [14] E. Indriati, N. Suharyani Azisa, E. Ivo Sihombing, and Z. Sukma Dewi Mokodompit, “Implementasi Algoritma K-Means Clustering Untuk Pengelompokan Status Penerima Kip Kuliah Mahasiswa Universitas Papua,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 6, 2024, doi: 10.36040/jati.v7i6.8222.
- [15] S. Paembonan and H. Abduh, “Penerapan Metode Silhouette Coefficient untuk Evaluasi Clustering Obat,” *PENA Tek. J. Ilm. Ilmu-Ilmu Tek.*, vol. 6, no. 2, 2021, doi: 10.51557/pt_jiit.v6i2.659.
- [16] J. W. Fernanda, “Pemodelan Persepsi Pembelajaran Online Menggunakan Latent Dirichlet Allocation,” *J. Stat. Univ. Muhammadiyah Semarang*, vol. 9, no. 2, 2021, doi: 10.26714/jsunimus.9.2.2021.79-85.
- [17] C. C. Silva, M. Galster, and F. Gilson, “Topic modeling in software engineering research,” *Empir. Softw. Eng.*, vol. 26, no. 6, 2021, doi: 10.1007/s10664-021-10026-0.
- [18] Akdeas Oktanae Widodo, F. Septiadi, and Nur Aini Rakhmawati, “Analisis Tren Konten Pada Vtuber Indonesia Menggunakan Latent Dirichlet Allocation,” *J. Inform. dan Rekayasa Elektron.*, vol. 6, no. 1, 2023, doi: 10.36595/jire.v6i1.718.
- [19] T. Titiana and D. H. Bangkalang, “Analisis Dan Penerapan Topic Modeling Pada Judul Tugas Akhir Mahasiswa Menggunakan Metode Latent Dirichlet Allocation (Lda),” *JIPi (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 8, no. 4, 2023, doi: 10.29100/jipi.v8i4.4254.
- [20] N. A. Rakhmawati, R. A. Zuhroh, Q. N. Muna, and V. R. Dianutami, “Klasterisasi Keyword Terkait Pornografi pada Media Sosial Twitter Menggunakan Latent Dirichlet Allocation,” *J. Inf. Eng. Educ. Technol.*, vol. 6, no. 2, 2022, doi: 10.26740/jieet.v6n2.p66-72.