

Klasifikasi Topik Pembahasan Mahasiswa ITS dalam Bermedia Sosial Menggunakan *Latent Dirichlet Allocation*

Muhammad Rifqi Hidayat¹, Naufal Firjatulloh Fano², Ribka Devina Margaretha³, Zada Alfarras Rasendriya⁴, Nur Aini Rakhmawati⁵

^{1,2,3,4,5} Departemen Sistem Informasi, Institut Teknologi Sepuluh Nopember

¹mrifqihdyt12@gmail.com

²naufal.fano@gmail.com

³ribkaadevina@gmail.com

⁴zadaalfarras@gmail.com

⁵nur.aini@is.its.ac.id

Media sosial sudah menjadi bagian dari kehidupan sehari-hari masyarakat segala kalangan di era digital ini. Di tengah maraknya penggunaan media sosial ini mahasiswa Institut Teknologi Sepuluh Nopember (ITS) ikut serta dalam berdiskusi di berbagai media sosial mengenai topik seputar kampus. Salah satu platform yang populer di kalangan mahasiswa ITS adalah Twitter. Pada platform ini terdapat sebuah akun media sosial yaitu @its_fess. @its_fess adalah akun Twitter yang dikelola oleh mahasiswa ITS untuk berbagai pemikiran, cerita, dan topik-topik yang berkaitan dengan kehidupan di ITS secara anonim. Dengan fenomena ini dibutuhkan suatu pemodelan topik yang mampu mengklasifikasi topik pembahasan mahasiswa ITS di media sosial, khususnya pada akun @its_fess. Pemodelan topik dilakukan menggunakan metode Latent Dirichlet Allocation (LDA), sebuah algoritma *text mining* untuk mengidentifikasi topik utama yang terdapat pada sebuah dokumen. Eksperimen pemodelan topik dengan metode LDA menyimpulkan bahwa jumlah topik yang terdapat dalam tweet @its_fess adalah 20 topik. Tweet dari akun @its_fess dikumpulkan menggunakan teknik scrapping. Hasil tweet tersebut direprocessing untuk selanjutnya dianalisis dengan metode LDA. Hasil pemodelan topik dievaluasi menggunakan perhitungan nilai koherensi. Evaluasi model mendapatkan nilai koherensi sebesar 0.518712 untuk jumlah topik sebanyak 20. Nilai ini menunjukkan bahwa model yang digunakan baik untuk menganalisis topik dalam tweet pada @its_fess.

Kata Kunci— Mahasiswa, Twitter, Pemodelan Topik, *Latent Dirichlet Allocation*, Nilai Koherensi

I. PENDAHULUAN

Media sosial telah menjadi bagian integral dalam kehidupan sehari-hari masyarakat segala kalangan di era digital ini. Media sosial memungkinkan penggunaannya untuk berbagi pandangan, pengalaman, dan pemikiran dengan cepat dan mudah. Tak terlupe juga kalangan mahasiswa [1]. Berbagai platform media sosial, seperti Instagram, Twitter, Facebook, dan lain sebagainya tidak hanya digunakan untuk berinteraksi sosial saja, tetapi juga sebagai saluran penting bagi mahasiswa untuk bertukar pikiran tentang berbagai pembahasan, termasuk pula isu-isu seputar perkuliahan dan perguruan tinggi mereka. Dengan dinamikanya yang khas, media sosial menciptakan lingkungan dimana topik-topik pembahasan yang beragam dapat berkembang.

Di tengah maraknya penggunaan media sosial, mahasiswa Institut Teknologi Sepuluh Nopember (ITS) turut ambil bagian untuk berdiskusi mengenai topik-topik seputar kampus di berbagai media sosial. Salah satu platform media sosial yang

sangat populer di kalangan mahasiswa ITS adalah Twitter. Pada platform tersebut, terdapat sebuah akun media sosial yang menjadi pusat perhatian bagi mahasiswa ITS, yaitu @its_fess. @its_fess adalah akun Twitter yang dikelola oleh mahasiswa ITS untuk berbagi pemikiran, cerita, pengalaman, dan topik-topik menarik yang berkaitan dengan kehidupan di kampus secara anonim.

Pentingnya memahami topik-topik pembahasan mahasiswa ITS dalam bermedia sosial seperti pada akun @its_fess tidak hanya untuk mengetahui tren serta isu yang sedang hangat di kalangan mahasiswa saja, tetapi juga memberikan wawasan yang dapat membantu pihak perguruan tinggi untuk menyusun strategi komunikasi dan interaksi yang lebih efektif di media sosial. Oleh karena itu, untuk mengetahui topik-topik pembahasan mahasiswa pada @its_fess, studi ini dilakukan dengan menerapkan teknik Latent Dirichlet Allocation (LDA) dalam mengklasifikasi topik pembahasan mahasiswa ITS di media sosial, khususnya di akun @its_fess. LDA adalah algoritma yang digunakan untuk mengidentifikasi topik-topik utama yang ada dalam sejumlah besar dokumen teks, yang dalam konteks ini adalah tweet-tweet yang diposting di akun @its_fess. Dengan menerapkan LDA, kita dapat menggali wawasan yang lebih dalam tentang topik-topik yang sedang dibahas oleh mahasiswa ITS di media sosial ini. Metode LDA yang diterapkan pada studi kasus ini menggunakan tools gensim dengan metode Variational Bayes.

Penelitian terdahulu yang berjudul “Analisis Topik Tagar Covidindonesia pada Instagram Menggunakan Latent Dirichlet Allocation”, telah menggunakan metode LDA untuk pengelompokan topik pada tweet dengan tagar Covidindonesia [2]. Selain itu, terdapat pula penelitian terdahulu berjudul “Analisis Tren Konten Pada Vtuber Indonesia menggunakan Latent Dirichlet Allocation”, yang juga menggunakan metode LDA untuk pengelompokan topik konten pada Vtuber Indonesia [3]. Selain itu, terdapat pula penelitian terdahulu yang berjudul “Analisis Topik Modelling Terhadap Penggunaan Sosial Media Twitter oleh Pejabat Negara” yang juga menggunakan metode LDA pada Twitter untuk pengelompokan topik pembahasan pejabat negara [4]. Terdapat pula penelitian berjudul “Analisis Topik Informasi Publik Media Sosial di Surabaya Menggunakan Pemodelan Latent Dirichlet Allocation (LDA)” yang menggunakan LDA untuk mengetahui topik informasi publik di Surabaya melalui radio Suara Surabaya [5]. Pada penelitian ini, tim peneliti menggunakan metode LDA untuk pengelompokan topik

pembahasan mahasiswa ITS pada akun @its_fess. Adapun penelitian ini bertujuan untuk menerapkan teknik LDA dalam mengklasifikasikan topik pembahasan yang dominan di antara mahasiswa ITS dalam media sosial Twitter, dengan fokus pada akun @its_fess sebagai studi kasus. Melalui penelitian ini, tim peneliti berharap dapat memberikan pemahaman lebih jauh mengenai topik-topik yang relevan bagi mahasiswa ITS dan bagaimana diskusi topik tersebut tercermin di media sosial.

II. TINJAUAN PUSTAKA

A. Pemodelan Topik

Pemodelan topik adalah konsep yang digunakan untuk *corpus* (kumpulan dokumen) di mana dokumen, yang terdiri dari banyak kata, dapat direpresentasikan ulang menjadi sebuah variabel dan angka [6]. Singkatnya pemodelan topik meringkas suatu dokumen menjadi sebuah *vector* yang terdiri dari perhitungan kata. Kumpulan kata ini selanjutnya dapat disimpulkan menjadi sebuah topik .

B. Gensim

Gensim merupakan salah satu *library* Python yang digunakan untuk pemodelan topik, pengindeksan dokumen, dan pengambilan kesamaan dari dokumen atau data yang besar. *Library* ini dibuat untuk dapat mengambil topik semantik dari sebuah data secara otomatis. Selain untuk pemodelan topik, toolkit yang bersifat *open source* ini juga dapat digunakan sebagai pemodelan ruang vektor dan dapat diimplementasikan menggunakan NumPy, SciPy, dan Cython secara opsional untuk meningkatkan kinerja dari Gensim. Cara kerja gensim adalah dengan menggunakan algoritma *TextRank* yang sangat sering digunakan, cara kerja ini didasarkan oleh kalimat pada teks yang diberi peringkat menggunakan algoritma *TextRank*. *TextRank* merupakan sebuah algoritma berbasis grafik untuk *natural language processing* (NLP). Algoritma ini berfungsi untuk membuat ringkasan dari suatu dokumen menjadi peringkat berbasis grafik [7].

C. Latent Dirichlet Allocation (LDA)

Latent Dirichlet Allocation (LDA) merupakan salah satu metode pengolahan *corpus* menjadi klasifikasi teks dan permodelan topik dengan model probabilistik generatif [8]. LDA juga dapat memproses data dalam jumlah besar dengan cara mengelompokkan dan menghubungkan yang menghasilkan luaran berupa daftar topik beserta dengan besar proporsi dari masing-masing topik. LDA bekerja dengan menjadikan kumpulan dokumen dan parameter sebagai input yang menghasilkan model berbobot yang dapat dinormalisasi menjadi probabilitas. Probabilitas sendiri terbagi menjadi dua jenis yaitu (a) probabilitas dokumen tertentu yang menghasilkan topik spesifik (b) probabilitas dari topik tertentu yang menghasilkan suatu kata dari kumpulan kosakata. Probabilitas kata yang muncul pada suatu berasal dari distribusi Dirichlet. Topik yang dihasilkan pada LDA merupakan hipotesis yang tidak tertulis pada dokumen [9]. LDA telah terbukti mampu untuk melakukan pemodelan topik berdasarkan data teks dalam bentuk *tweet* dengan bahasa Indonesia [10]. Penelitian lain yang dilakukan oleh Oktaviana, dkk menggunakan metode LDA untuk melakukan permodelan topik dari artikel suatu portal berita menunjukkan bahwa LDA

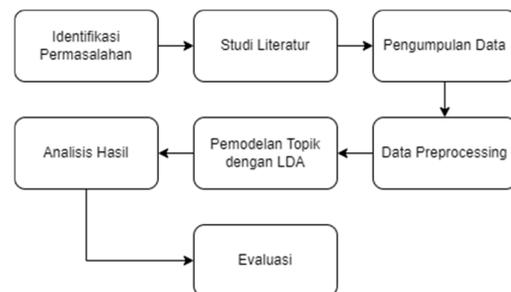
menghasilkan 8 topik dengan 2 topik tidak mudah diinterpretasikan [11].

D. Akun @its_fess

Twitter merupakan salah satu platform media sosial yang banyak digunakan oleh masyarakat untuk mencari informasi terbaru maupun saling bertukar pesan dengan sesama penggunanya. Twitter tergolong sebagai layanan mikroblog yang memungkinkan penggunanya untuk membaca postingan (*tweet*) dari pengguna lain dalam bentuk teks yang dapat ditambahkan media berupa gambar maupun video [12]. Salah satu akun Twitter yang paling sering digunakan oleh mahasiswa ITS dalam bertukar informasi adalah akun @its_fess. Pada akun ini mahasiswa yang merupakan pengguna Twitter dapat bertukar informasi dalam bentuk *tweet* yang bersifat anonim. Tercatat pada tanggal 20 September 2023 akun @its_fess telah diikuti oleh 23.300 pengguna.

III. METODOLOGI

Bagian ini akan menjelaskan mengenai metode penelitian yang digunakan serta tahapan-tahapan yang dilakukan dalam penelitian ini. Alur penelitian yang dilakukan digambarkan dalam diagram berikut.



Gbr. 1 Tahapan Metodologi Penelitian

A. Identifikasi Permasalahan

Identifikasi permasalahan dilakukan untuk menentukan tujuan dari dilaksanakannya penelitian ini. Proses identifikasi permasalahan melibatkan pendalaman awal mengenai latar belakang penelitian, analisis data awal dari @its_fess, dan penentuan pendekatan *gensim* untuk analisis LDA. Penelitian ini dilakukan dengan tujuan untuk mengetahui topik-topik yang menjadi pembahasan mahasiswa ITS dalam bermedia sosial. Dalam studi kasus ini, tim peneliti menggunakan akun Twitter @its_fess sebagai objek penelitian. Adapun metode yang digunakan adalah *Latent Dirichlet Allocation* (LDA) dengan menggunakan tools *gensim* dengan metode *Variational Bayes*. Berdasarkan pemahaman ini, hipotesis awal dirumuskan untuk memahami bagaimana LDA membantu mengklasifikasi topik-topik pembahasan mahasiswa ITS yang dominan pada studi kasus tersebut.

B. Studi Literatur

Dalam tahapan studi literatur, proses dimulai dengan mendalami konsep dasar metode yang digunakan, yaitu LDA, termasuk metode yang digunakan, penerapannya, dan studi terdahulu yang telah menerapkan LDA dalam konteks serupa. Metode pengambilan data untuk objek penelitian juga dipelajari pada tahapan ini dengan berkaca pada penelitian terdahulu pula. Tahapan ini juga melibatkan proses eksplorasi terhadap studi

penggunaan media sosial oleh mahasiswa. Selain itu, tahapan ini mencakup pula analisis literatur terhadap akun Twitter @its_fess sebagai studi kasus di penelitian ini. Adanya tahapan ini memberikan dasar teoritis yang kuat untuk melaksanakan penelitian.

C. Pengumpulan Data

Tahapan ini mencakup data crawling atau mengumpulkan data cuitan pada akun Twitter @its_fess. Data diperoleh dengan menggunakan proses data scraping. Proses tersebut dapat digunakan untuk mengimpor data cuitan pada akun Twitter @its_fess menjadi spreadsheet atau JSON. Metode scraping dilakukan dengan bahasa pemrograman Python. Dari tahapan ini, dihasilkan dataset berisikan data cuitan mentah yang kemudian diunggah ke akun Zenodo [13].

D. Data Preprocessing

Data yang telah dikumpulkan kemudian akan melalui proses cleaning pada tahapan ini. Proses ini mencakup penghapusan duplikasi data, penghapusan tanda baca dan karakter yang tidak diperlukan, serta pembuangan stopwords, seperti di, yang, dari, dan lain sebagainya [14]. Selain itu, data mentah yang diambil dari proses scraping juga akan diambil hanya bagian kolom twitnya saja. Kemudian, data akan melalui proses tokenisasi yaitu pemecahan menjadi token-token kata tunggal yang akan memudahkan proses pengelompokan topik dengan LDA nantinya. Setelah data telah melalui semua proses ini, artinya data sudah siap untuk diolah dengan metode LDA di tahapan berikutnya. Dataset dari hasil tahapan ini diunggah pula pada akun Zenodo [13].

E. Pemodelan Topik dengan LDA

Pada tahapan pemodelan topik, tim peneliti melakukan pemodelan dari token-token yang sudah dikumpulkan dengan metode Latent Dirichlet Allocation (LDA). Tahapan ini dilakukan untuk mendeteksi topik-topik yang muncul pada token dan proporsi seberapa banyak topik itu muncul. Pemodelan topik akan menggunakan tools gensim, yang merupakan salah satu library dalam python [15].

F. Analisis Hasil

Pada tahap analisis data, tim peneliti memanfaatkan dataset teks yang dihasilkan melalui penerapan Metode Latent Dirichlet Allocation (LDA) pada studi kasus @its_fess dalam penelitian berjudul "Klasifikasi Topik Pembahasan Mahasiswa ITS dalam Bermedia Sosial Menggunakan Metode Latent Dirichlet Allocation (Studi kasus: @its_fess)." Data tersebut adalah hasil dari proses LDA yang telah diubah menjadi kelompok-kelompok topik berdasarkan kesamaan kata kunci atau pola dalam teks. Tim peneliti kemudian melakukan analisis visual terhadap data ini dengan menggambarkan visualisasi perbandingan antar topik-topik yang telah diidentifikasi melalui grafik bar-chart.

G. Evaluasi

Setelah hasil telah didapatkan, selanjutnya dilakukan evaluasi untuk menguji hasil yang didapatkan. Evaluasi dilakukan dengan menggunakan uji koherensi. Uji koherensi (coherence test) adalah metrik yang digunakan untuk mengevaluasi performa model Latent Dirichlet Allocation (LDA) dalam pemodelan tema (topic modelling) pada teks.

Evaluasi dengan uji koherensi adalah metrik evaluasi yang mengukur nilai suatu topik dengan membandingkan kesamaan semantik dengan kata-kata pada topik. Semakin tinggi nilai dari skor koherens, menandakan semakin baik model yang digunakan dalam penelitian.

IV. HASIL DAN PEMBAHASAN

A. Hasil Pengumpulan Data

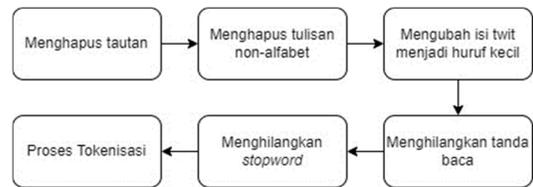
Pengumpulan data dilakukan dengan melakukan scraping data dari website Twitter dengan menggunakan bahasa pemrograman Python. Data twit yang digunakan diambil dari akun @its_fess pada rentang waktu 2023, lebih spesifiknya pada bulan Januari hingga Agustus 2023. Dari hasil scraping, didapatkan data dengan fitur indeks, tanggal twit, isi twit, jumlah quote, jumlah retweet, jumlah favorite, bahasa, indeks akun, nama akun, dan juga tautan twit. Data yang didapatkan dihimpun dalam bentuk tabel yang disimpan dalam file bertipe .csv. Data yang didapatkan masih berupa data mentah, yang kemudian akan melalui tahapan pre-processing data.

created_at	id_str	full_text	quote_cou	reply_cou	retweet_c	favorite_c
Sat Apr 29 23:56	1,65E+18	(Titipan) mas/mbak its,	0	7	0	0
Sat Apr 29 23:47	1,65E+18	Its/ Info kost cowo peri	0	0	0	0
Sat Apr 29 22:30	1,65E+18	Its/ kalau dari stasiun g	0	2	0	0
Sat Apr 29 20:25	1,65E+18	Its/ jam berapa ya kira	0	1	0	1
Sat Apr 29 18:20	1,65E+18	in sender diamond kah	0	1	0	0
Sat Apr 29 18:20	1,65E+18	rek, its/ siapapun yg m	1	1	0	0
Sat Apr 29 17:44	1,65E+18	Its/ koko sipil 21 nrp 91	3	0	0	4
Sat Apr 29 17:07	1,65E+18	Its/ test	0	3	0	0
Sat Apr 29 16:45	1,65E+18	ada yg open anjem dar	0	2	0	1
Sat Apr 29 16:41	1,65E+18	Its/ ini base beneran m	0	0	0	0
Sat Apr 29 16:25	1,65E+18	Its/ nrp 132 mindep ula	1	0	0	0
Sat Apr 29 16:13	1,65E+18	Its/ ada yang pinjem di	0	1	0	0
Sat Apr 29 16:07	1,65E+18	Its/ rek yg punya info ki	0	0	0	2
Sat Apr 29 15:50	1,65E+18	Its/ cowok2 di its ada g	3	13	0	13
Sat Apr 29 15:48	1,65E+18	ada yang ngejual rice cc	0	0	0	0
Sat Apr 29 15:38	1,65E+18	Rek, di Deket its/ atau c	0	3	0	1
Sat Apr 29 15:00	1,65E+18	Its/fess beneran off? Tf	0	0	0	15

Gbr. 2 Contoh data hasil *scrapping* pada akun Twitter @its_fess

B. Hasil Preprocessing Data

Data mentah yang didapatkan dari hasil scraping kemudian diolah melalui tahapan preprocessing data yang menjadikan data siap diolah. Langkah-langkah yang dilalui pada tahapan ini digambarkan pada diagram berikut.



Gbr. 3 Alur tahapan *preprocessing data*

Untuk mengubah data mentah menjadi data yang siap dimodelkan dengan metode LDA, langkah-langkah pada Gambar 3 perlu untuk dilakukan. Adapun stopwords yang dimaksud pada gambar di atas adalah kata-kata yang umumnya diabaikan dalam analisis teks karena tidak memiliki konteks yang cukup berarti terhadap hasil pemodelan. Dalam penelitian ini, digunakan library stopwords NLTK beserta dengan tambahan beberapa stopwords yang seperti "arek", "rek", "mas", dll. Setelah melakukan penghapusan untuk beberapa item yang tidak diperlukan dalam analisis dan mengubah seluruh item menjadi huruf kecil, twit dipecah menjadi beberapa token kata tunggal. Berikut adalah beberapa hasil akhir dari tahapan preprocessing.

```

                                tokenized_text
0      [ok, close, buka, kl, basenya, udah, aktif]
1      [open, fallback, itsfess, come, back, tweet, r...
2      [base, off, yaaaaa]
3      [udah, banget, berkecimpung, food, kah, tanyat...
4      [info, perpanjangan, kuesioner, ipd]
...
4952      []
4953      []
4954      [liburan, ngapain, sih, turu]
4955      [gais, nanya]
4956      [sender, maba]
    
```

Gbr. 4 Hasil tahapan *preprocessing*

C. Hasil Pemodelan Topik dengan LDA

Dalam melakukan pemodelan data yang telah melalui proses pre-processing digunakan sebagai input. Pemodelan dilakukan dengan menetapkan variable jumlah topik yang didapat melalui uji koherensi. Pada penelitian ini didapat bahwa nilai koherensi tertinggi ada pada 20 topik. Luaran dari pemodelan topik akan menghasilkan topik, probabilitas kata, dan top word. Dari luaran tersebut dapat diketahui kata apa yang sering muncul pada suatu topik beserta dengan nilai probabilitas dari kata tersebut. Berikut merupakan luaran dari pemodelan topik pada penelitian ini dengan num_topics = 20 dan 10 top word pada setiap topik.

TABEL I
HASIL PEMODELAN TOPIK

Topik	Top Word
0	0.056*"review" + 0.053*"pkm" + 0.026*"smt" + 0.021*"dosen" + 0.015*"wajib" + 0.014*"itu" + 0.012*"dosennya" + 0.011*"kapan" + 0.009*"word" + 0.009*"sih"
1	0.019*"spill" + 0.015*"kira kira" + 0.012*"mulu" + 0.012*"kira" + 0.011*"tiket kereta" + 0.010*"di" + 0.010*"pernah" + 0.010*"reply" + 0.010*"aja" + 0.010*"ngasih"
2	0.021*"menurut kalian" + 0.021*"worth_it" + 0.015*"teknik" + 0.014*"kuliah" + 0.013*"sini" + 0.012*"di_sini" + 0.012*"di" + 0.011*"hal" + 0.011*"tia" + 0.011*"it"
3	0.045*"sks" + 0.022*"yaa" + 0.017*"maaf" + 0.016*"organisasi" + 0.014*"tau tempat" + 0.013*"lebih" + 0.012*"mohon" + 0.011*"rekkk" + 0.011*"sendiri" + 0.011*"rencana"
4	0.079*"mau tanya" + 0.050*"skem" + 0.042*"tanya" + 0.025*"yuk" + 0.022*"ukt" + 0.020*"mau" + 0.014*"ets" + 0.014*"di" + 0.013*"info_kos" + 0.012*"bayar"
5	0.038*"maba" + 0.021*"join" + 0.020*"bikin" + 0.017*"matkul" + 0.016*"udah" + 0.014*"sender" + 0.014*"ayang" + 0.013*"belum" + 0.013*"nambah" + 0.013*"jadwal"
6	0.028*"balik ke" + 0.025*"balik" + 0.024*"halo" + 0.023*"tanggal berapa" + 0.022*"tanggal" + 0.017*"aku" + 0.016*"berapa" + 0.015*"kamar_mandi" + 0.015*"wa" + 0.013*"ges"
7	0.031*"ambil" + 0.020*"murah" + 0.017*"di" + 0.017*"siakad" + 0.017*"info tempat" + 0.013*"beli" + 0.013*"sekitaran" + 0.012*"dimana" + 0.011*"info" + 0.011*"tempat"
8	0.056*"boleh minta" + 0.035*"info" + 0.019*"kemarin" + 0.018*"dosen" + 0.015*"minta" + 0.015*"mbakmas" + 0.013*"ku" + 0.013*"boleh" + 0.011*"mutualan" + 0.011*"buka"

9	0.034*"reviewnya" + 0.031*"war" + 0.024*"mau" + 0.019*"ngulang" + 0.016*"mau_nanya" + 0.016*"matkul_pengayaan" + 0.015*"help" + 0.015*"k" + 0.015*"kalian" + 0.015*"matkul"
10	0.055*"mau_nanya" + 0.049*"permisi" + 0.030*"minta_tolong" + 0.027*"nanya" + 0.021*"nanti" + 0.020*"anak" + 0.017*"mau" + 0.016*"sender" + 0.014*"hehe" + 0.014*"dm"
11	0.027*"tips" + 0.023*"bf" + 0.018*"haloo" + 0.018*"ta" + 0.016*"tedx" + 0.016*"ngajuin" + 0.015*"rep" + 0.014*"mulai" + 0.014*"random" + 0.014*"akun"
12	0.024*"misi" + 0.020*"masmbak" + 0.017*"itu" + 0.017*"extend" + 0.016*"kelas" + 0.015*"buat" + 0.013*"info" + 0.013*"info_kehilangan" + 0.013*"kan" + 0.012*"mending"
13	0.016*"iisma" + 0.016*"olahraga" + 0.016*"dari" + 0.015*"mutualan" + 0.015*"wts" + 0.013*"seru" + 0.013*"luar" + 0.012*"kalian" + 0.012*"arsi" + 0.011*"mulai"
14	0.028*"oprec" + 0.021*"ukm" + 0.020*"sem" + 0.019*"kok" + 0.019*"tutorial" + 0.018*"uang" + 0.017*"libur" + 0.016*"baru" + 0.015*"bgt" + 0.013*"aku"
15	0.039*"matkul_pengayaan" + 0.031*"pengayaan" + 0.021*"boleh" + 0.020*"ikut" + 0.020*"banget" + 0.017*"wdyt" + 0.014*"misal" + 0.014*"ini" + 0.013*"serius" + 0.013*"hima"
16	0.034*"jam_berapa" + 0.034*"hari_ini" + 0.024*"ini" + 0.022*"liburan" + 0.017*"hari" + 0.016*"cara" + 0.015*"tim" + 0.014*"berapa" + 0.013*"buat" + 0.013*"siapa"
17	0.020*"di" + 0.019*"kegiatan" + 0.019*"tolong" + 0.018*"atm" + 0.017*"info" + 0.017*"yaynay" + 0.017*"gebang" + 0.016*"buka_jasa" + 0.016*"isi" + 0.014*"infonya"
18	0.030*"dimana_yaa" + 0.026*"kimia" + 0.024*"yay_or" + 0.021*"masmba" + 0.018*"or" + 0.018*"inpo" + 0.015*"pgn" + 0.014*"biasanya" + 0.013*"hujan" + 0.012*"dimana"
19	0.075*"info_kos" + 0.032*"kos" + 0.019*"kalian" + 0.017*"info" + 0.016*"dicari" + 0.014*"dept" + 0.013*"daerah_keputih" + 0.013*"cewek" + 0.012*"foto" + 0.012*"sedih"

Berdasarkan tabel diatas dapat dilakukan analisis luaran pemodelan topik berdasarkan top word pada masing-masing topik untuk mengidentifikasi pembahasan mahasiswa ITS dalam bermedia sosial. Berikut merupakan hasil analisis pada setiap topik.

TABEL III
ANALISIS LUARAN PEMODELAN TOPIK

Topik	Pembahasan
0	Pembicaraan terkait dosen ITS
1	Pertanyaan mengenai aktivitas sehari-hari mahasiswa
2	Pendapat terkait kepuasan kuliah teknik di ITS
3	Saran dan pandangan terhadap perilaku organisasi
4	Pertanyaan terkait informasi administrasi dan akademik ITS
5	Jadwal dan jumlah mata kuliah pada mahasiswa baru
6	Informasi kembali menuju kota asal pada saat liburan
7	Informasi tempat belanja di sekitar ITS
8	Pertanyaan yang ditujukan kepada mahasiswa senior
9	Informasi mengenai proses pengambilan matakuliah pada masa pengisian Formulir Rencana Studi (FRS)
10	Pertanyaan mengenai informasi spesifik

11	Tidak dapat diinterpretasikan
12	Informasi kehilangan barang mahasiswa
13	Tidak dapat diinterpretasikan
14	Informasi pendaftaran Unit Kegiatan Mahasiswa
15	Pendapat seputar mata kuliah pengayaan
16	Tidak dapat diinterpretasikan
17	Informasi lokasi penunjang kebutuhan mahasiswa ITS
18	Pertanyaan mengenai lokasi
19	Informasi terkait lokasi kos di sekitar lingkungan ITS

Semakin besar lingkaran menandakan bahwa topik tersebut sering muncul menjadi topik pembahasan. Sementara itu, grafik bar pada gambar 6 menggambarkan kata-kata yang dominan muncul pada topik. Terdapat 30 kata atau terms yang paling sering disebutkan, yaitu “mau tanya”, “info kos”, “review”, “pkm”, “boleh minta”, “mau nanya”, “skem”, “permisi”, “sks”, “matkul pengayaan”, “tanya”, “reviewnya”, “jam berapa”, “ambil”, “war”, “pengayaan”, “hari ini”, “dosen”, “smt”, “dimana yaa”, “minta tolong”, “maba”, “balik ke”, “kos”, “oprec”, “nanya”, “kimia”, “yuk”, “balik”, dan “matkul”. Ketika panah diarahkan pada kata, lingkaran akan berubah sesuai seberapa sering kata tersebut muncul pada topik.

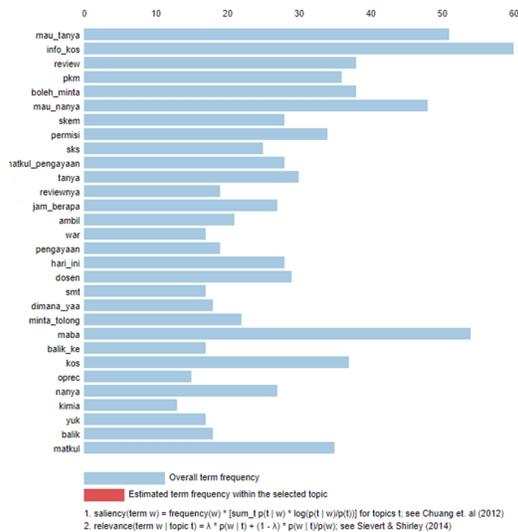
D. Visualisasi Pemodelan Topik

Dengan menggunakan library gensim dan pyLDAvis, dihasilkan visualisasi dari pemodelan topik yang telah dilakukan. Visualisasi tersebut dapat dilihat pada gambar di bawah ini



Gbr. 5 Visualisasi Pemodelan Topik

Kata-kata yang dominan muncul pada topik tersebut digambarkan pada grafik bar berikut.



Gbr. 6 Visualisasi kata dominan pada topik

Pada Gambar 5, lingkaran di sisi kiri menggambarkan topik-topik dari twit yang telah diklasifikasikan. Besar kecil lingkaran menggambarkan seberapa sering topik tersebut muncul.

E. Hasil Evaluasi

Evaluasi dengan menggunakan uji koherensi dilakukan untuk memastikan apakah topik modelling yang dilakukan sudah merupakan hasil pemodelan yang paling baik. Parameter yang dijadikan acuan pada uji koherensi adalah jumlah topik yang digunakan untuk pemodelan. Tabel nilai koherensi dari topik yang dihasilkan dalam penelitian ini dapat dilihat pada tabel di bawah ini.

TABEL IIIII
NILAI EVALUASI KOHERENSI

Jumlah Topik	Nilai Koherensi
1	0.3495
2	0.332655
3	0.310978
4	0.354352
5	0.353190
6	0.382328
7	0.393783
8	0.409657
9	0.425487
10	0.438001
11	0.432393
12	0.464410
13	0.475237
14	0.497327
15	0.492226
16	0.513789
17	0.504349
18	0.498685
19	0.518609
20	0.518712

Semakin besar nilai koherensi, menandakan semakin baik model yang dipakai. Dari tabel di atas, nilai koherensi terbaik pada penelitian ini ada pada jumlah topik sebanyak 20 dengan nilai koherensi sebesar 0.518712.

V. KESIMPULAN

Kesimpulan penelitian ini didapatkan berdasarkan data twit yang diambil dari akun @its_fess pada rentang waktu Januari hingga Agustus 2023 dengan melakukan metode pemodelan topik menggunakan Latent Dirichel Allocation (LDA). Metode yang telah dilakukan menghasilkan 20 topik berdasarkan probabilitas kata tersebut muncul dalam sebuah twit. Dari 20 topik tersebut masih terdapat 3 topik yang tidak dapat diinterpretasikan. Berdasarkan evaluasi menggunakan nilai koherensi, ditemukan nilai koherensi terbaik pada penelitian ini sebesar 0.518712 pada jumlah topik sebanyak 20.

REFERENSI

- [1] V. Nainggolan, S. A. Randonuwu, and G. J. Waleleng, "Peranan Media Sosial Instagram dalam Interaksi Sosial Antar Mahasiswa Fakultas Ilmu Sosial dan Politik Unsrat Manado," *J. Acta Diurna*, vol. 7, no. 4, pp. 1–15, 2018, [Online]. Available: <https://ejournal.unsrat.ac.id/index.php/actadiurnakomunikasi/article/view/22022>.
- [2] K. R. A. P. Santoso, A. Husna, N. W. Putri, and N. A. Rakhmawati, "Analisis Topik Tagar Covidindonesia pada Instagram Menggunakan Latent Dirichlet Allocation," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 7, no. 1, pp. 1–9, 2022, doi: 10.14421/jiska.2022.7.1.1-9.
- [3] 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, pp. 56–63, 2023, doi: 10.36595/jire.v6i1.718.
- [4] P. Patmawati and M. Yusuf, "Analisis Topik Modelling Terhadap Penggunaan Sosial Media Twitter oleh Pejabat Negara," *Build. Informatics, Technol. Sci.*, vol. 3, no. 3, pp. 122–129, 2021, doi: 10.47065/bits.v3i3.1012.
- [5] K. B. Putra and R. P. Kusumawardani, "Analisis Topik Informasi Publik Media Sosial di Surabaya Menggunakan Pemodelan Latent Dirichlet Allocation (LDA)," *J. Tek. ITS*, vol. 6, no. 2, pp. 4–9, 2017, doi: 10.12962/j23373539.v6i2.23205.
- [6] H. Zhao, D. Phung, V. Huynh, Y. Jin, L. Du, and W. Buntine, "Topic Modelling Meets Deep Neural Networks: A Survey," *IJCAI Int. Jt. Conf. Artif. Intell.*, pp. 4713–4720, 2021, doi: 10.24963/ijcai.2021/638.
- [7] S. Sareen, "Text Summarisation with Gensim (TextRank Algorithm)," 2018. [https://medium.com/@shivangisareen/text-summarisation-with-gensim-textrank-46bbb3401289#:~:text=Text Summarisation with Gensim \(TextRank algorithm\)-&text=This summarising is based on,based ranking algorithm for NLP.\(accessed Sep. 20, 2023\)](https://medium.com/@shivangisareen/text-summarisation-with-gensim-textrank-46bbb3401289#:~:text=Text Summarisation with Gensim (TextRank algorithm)-&text=This summarising is based on,based ranking algorithm for NLP.(accessed Sep. 20, 2023)).
- [8] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *J. Mach. Learn. Res.*, vol. 3, no. null, pp. 993–1022, Mar. 2003.
- [9] J. C. Campbell, A. Hindle, and E. Stroulia, "Latent Dirichlet Allocation: Extracting Topics from Software Engineering Data," *Art Sci. Anal. Softw. Data*, pp. 139–159, 2015, doi: 10.1016/B978-0-12-411519-4.00006-9.
- [10] E. S. Negara, D. Triadi, and R. Andryani, "Topic Modelling Twitter Data with Latent Dirichlet Allocation Method," *ICECOS 2019 - 3rd Int. Conf. Electr. Eng. Comput. Sci. Proceeding*, pp. 386–390, 2019, doi: 10.1109/ICECOS47637.2019.8984523.
- [11] A. K. N. Oktaviana, N. A. S. ER, I. B. M. Mahendra, I. G. S. Astawa, I. G. A. Wibawa, and I. K. A. Mogi, "Pemodelan Topik Artikel Berita Menggunakan Structural Topic Model dan Latent Dirichlet Allocation," *JELIKU (Jurnal Elektron. Ilmu Komput. Udayana)*, vol. 11, no. 3, p. 469, 2022, doi: 10.24843/jlk.2023.v11.i03.p02.
- [12] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, a Social Network or a News Media?," in *Proceedings of the 19th International Conference on World Wide Web*, 2010, pp. 591–600, doi: 10.1145/1772690.1772751.
- [13] M. R. Hidayat, N. F. Fano, R. D. Margaretha, Z. A. Rasendriya, and N. A. Rakhmawati, "naufalfano/itsfess_topicmodelling_LDA: ItsFess LDA Topic Modelling Dataset." Zenodo, Oct. 2023, doi: 10.5281/zenodo.8400008.
- [14] F. F. Rachman and S. Pramana, "Analisis Sentimen Pro dan Kontra Masyarakat Indonesia tentang Vaksin COVID-19 pada Media Sosial Twitter," *Heal. Inf. Manag. J.*, vol. 8, no. 2, pp. 100–109, 2020, [Online]. Available: <https://inohim.esaunggul.ac.id/index.php/INO/article/view/223/175>.
- [15] Khairul Hudha Nasution, Widodo, and Bambang Prasetya Adhi, "Sistem Deteksi Topik Politik Pada Twitter Menggunakan Algoritma Latent Dirichlet Allocation," *PINTER J. Pendidik. Tek. Inform. dan Komput.*, vol. 5, no. 1, pp. 76–83, 2021, doi: 10.21009/pinter.5.1.10.