# Pemodelan Topik pada Cuitan tentang Penyakit Tropis di Indonesia dengan Metode *Latent Dirichlet Allocation*

Dziky Ridhwanullah<sup>1\*)</sup>, Dhomas Hatta Fudholi<sup>2)</sup>

<sup>1)</sup> Magister Informatika, Universitas Islam Indonesia

<sup>2)</sup> Informatika, Universitas Islam Indonesia

<sup>1)</sup> dzikyr@students.uii.ac.id, <sup>2)</sup> hatta.fudholi@uii.ac.id

#### **ABSTRACT**

Indonesia has a wide area and society. Therefore, a lot of information appear through social media, especially Twitter. This study aims to find out about conversation topics discussed by Indonesian people related to tropical diseases especially leprosy, malaria, and dengue fever. To find out the discussion topics, it can use the modeling topics analysis. One of the methods in topic modeling is Latent Dirichlet Allocation (LDA). Tweet data on tropical diseases in Indonesia was analyzed through this method. The study results showed that LDA was succeed in modeling the trend of Indonesian people's conversation topics related to tropical diseases. It obtained as many as 5 topics with a coherence value of 0.576453. Based on the results of the topic modeling, it can be concluded that the topics are such as the used funds to eradicate malaria and dengue fever, covid-19, blindness and leprosy, and its treatments and preventions.

Keywords: Twitter, Topic modeling, Latent Dirichlet Allocation

## I. PENDAHULUAN

Kemajuan era digital dapat menjadikan jalinan komunikasi antar pihak semakin cepat dan mudah, salah satu contohnya melalui media sosial. Platform media sosial digunakan sebagai komunikasi oleh masyarakat luas. Media sosial dapat digunakan untuk menciptakan kesadaran akan kesehatan seperti Twitter. Media sosial Twitter semakin banyak digunakan untuk diskusi (Grover, Kumar, & Davies, 2018), mengumpulkan informasi (Naveen Pemmaraju, Michael A. Thompson, 2017), penyebaran pengetahuan (Mallett et al., 2019) baik dalam praktis medis maupun penelitian ilmiah. Sehingga Twitter disebut dengan sumber data *real time* yang kaya akan pengetahuan terutama pada kesehatan.

Penyakit tropis biasanya menyerang suatu daerah yang memiliki iklim tropis. Negara Indonesia merupakan salah satu negara yang beriklim tropis sehingga memiliki peluang besar dihadapkan dengan penyakit ini. Penyakit tropis dapat disebabkan oleh bakteri, virus dan parasit (Alamudi, 2018). Penyakit tropis seperti malaria, demam berdarah, diare, tetanus, diphteri, amubiasis, demam chikungunya, ascariasis (Soegijanto, 2016) dan lain sebagainya. Berdasarkan pusdatin tahun 2019 menyatakan kasus penyakit tropis yang sering terjadi di Indonesia tercatat kasus kusta mencapai 17.439, malaria sebesar 256.680.000 dan demam berdarah 138.127 (Kesehatan & Indonesia, 2019).

Indonesia memiliki daerah yang sangat luar dan jumlah penduduk yang sangat besar, sehingga banyak informasi muncul melalui media sosial khususnya Twitter. Untuk mengetahui tren topik pembicaraan masyarakat Indonesia terkait dengan kesehatan yakni penyakit tropis dapat menggunakan analisis topik modeling. Pemodelan topik merupakan teknik yang digunakan dalam pendekatan *text mining* dan *text analysis* dalam menemukan data teks yang tersembunyi dan hubungan antar teks yang saling berkaitan dari suatu korpus (Jelodar & Wang, 2018). Salah satu metode dalam pemodelan topik adalah menggunakan *Latent Dirichlet Allocation* (LDA). Kelebihan LDA adalah menggunakan model probabilistik dengan topik yang dapat diinterpretasikan. LDA dalam penelitian sebelumnya digunakan untuk menganalisis topik terhadap pesan sosial media

(Putra, 2017), jurnal penelitian kesehatan (Yoga Sahria & Fudholi, 2021) dan biokimia (Kang & Kim, 2019).

Informasi mengenai penyakit tropis dapat dianalisis dan menghasilkan wawasan yang akan memberikan nilai serta keuntungan penggunaan LDA dalam menemukan hubungan antar teks yang saling berkaitan. Maka penelitian ini mengusulkan membuat pemodelan topik dengan LDA terhadap penyakit tropis yakni kusta, malaria dan demam berdarah dengan media Twitter. Metodenya adalah mencari nilai koherensi terlebih dahulu untuk menentukan jumlah topik yang terbaik. Kemudian dilakukan analisis secara kualitatif berdasarkan hasil tiap topiknya. Sehingga harapanya dari hasil analisis dapat digunakan pemerintah dan pelayanan kesehatan memberikan informasi maupan gerakan dalam upaya pencegahan atau penyebaran terjadinya kasus penyakit tersebut.

# II. TINJAUAN PUSTAKA

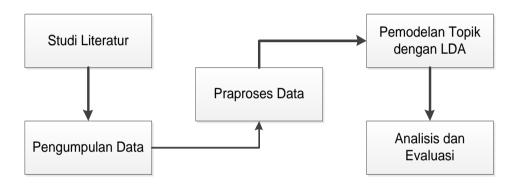
Kemajuan teknologi informasi menghadirkan berbagai media sosial yang dapat digunakan masyarakat luas sebagai media komunikasi. Melalui media sosial, masyarakat dapat menyebarkan informasi, memberikan saran ataupun kritikan, mengungkapkan emosional yang sedang dirasakan. Sehingga media sosial menghasilkan data informasi yang sangat besar atau disebut dengan big data. Agar dapat mengambil informasi yang penting dalam data tersebut dapat menggunakan analisis teks. Ada banyak hal yang dapat dilakukan oleh analisis teks seperti spam detection, news articles categorization, social media analysis and monitoring, bio-medical, security intelligence, marketing and crm, sentiment analysis, ad placements, virtual assistants dan chatbots (Sarkar, 2009) Penelitian sebelumnya analisis teks dalam bidang kesehatan digunakan untuk mengamati sentimen masyarakat melalui media sosial Twitter terhadap opini mengenai covid-19. Hasil penelitian tersebut 79% polaritas tweet negatif, 11% netral dan 10% positif (Kurniawan & Apriliani, 2020). Kemudian pada penelitian berikutnya, pembuatan aplikasi *chatbot* pada media sosial Line oleh pelayanan publik kesehatan di kota Malang guna mempermudah komunikasi mengenai kesehatan dengan masyarakat (Hormansyah & Aulia, 2018) dan pada penelitian berikutnya chatbot juga diterapkan pada layanan kesehatan yakni puskesmas dan hasil studi menyatakan tingkat kepuasan pengguna terhadap adanya *chatbot* pada layanan kesehatan sangat puas yakni sebesar 76.7% (Falah & Syamsidar, 2021).

Pemodelan topik merupakan suatu algoritma yang digunakan untuk menemukan topik dari kumpulan teks yang besar dan tidak terstruktur dan juga merupakan model yang menyimpulkan topik dengan mengelompokkan kata-kata dengan arti serupa. Karena fitur ini, pemodelan topik banyak digunakan untuk menganalisis topik dan tren. Penelitian sebelumnya informasi atau pesan yang disampaikan oleh masyarakat luas kepada radio dan diteruskan ke media sosial baik Twitter dan Facebook yang kemudian dari banyaknya informasi yang ada akan diklasifikasikan oleh LDA ke dalam topik-topik yang muncul dari hasil pemodelan. Hasil dalam studi ini menghasilkan 10 passes terbaik berdasarkan nilai perplexity dan kestabilan dapat dicapai pada iterasi kesepuluh setiap distribusi topik yang terpilih secara acak dan terbentuk 4 topik berdasarkan analisis nilai perplexity (Putra, 2017). Kemudian pada penelitian berikutnya berbagai judul penelitian di bidang kesehatan dari berbagai jurnal SINTA akan dikumpulkan dan untuk mengetahui bagaimana trend judul penelitian kesehatan maka pada studi ini menggunakan metode pemodelan topik LDA. Hasilnya LDA dapat melakukan pemodelan topik terhadap judul penelitian di bidang penelitian kesehatan di Indonesia. (Yoga Sahria & Fudholi, 2021). Selain itu juga diterapkan pada penelitian-penelitian mengenai biokimia. Hal ini dikarenakan studi yang berkembang pesat mengakibatkan penelitian biokimia tersebar luas yang mengakibatkan sulitnya membedakan bidang-bidang yang terkait lainnya. Agar dapat mengidentifikasi topik penelitian di bidang biokimia selama dua puluh tahun terakhir dan menganalisis perubahan trennya secara kuantitatif maka pada studi ini menerapkan metode pemodelan topik LDA (Kang & Kim, 2019).

Perbedaan penelitian sebelumnya dengan penelitian ini adalah media Twitter digunakan sebagai sumber data dimana dilakukan pengambilan data secara *streaming*. Kemudian dilakukan analisis pemodelan topik menggunakan LDA terhadap penyakit tropis di Indonesia. Adapun fokus penyakitnya seperti kusta, malaria dan demam berdarah. Sehingga dari hasil analisis dapat digunakan untuk pemerintah dan pelayanan kesehatan memberikan informasi maupan gerakan dalam upaya pencegahan atau penyebaran terjadinya kasus penyakit tersebut.

#### III. METODE PENELITIAN

Metode penelitian ini memiliki beberapa tahap yang harus dilalui seperti studi literatur, pengumpulan data, praproses data, pemodelan topik dengan LDA, analisis dan evaluasi seperti yang ditunjukkan pada Gambar 1.



Gambar 1 Metode Penelitian

#### 3.1. Studi Literatur

Studi literatur dilakukan dengan mempelajari keilmuan yang berhubungan dengan topik penelitian yakni penyakit tropis, media sosial Twitter, *text mining, text analytic* dan pemodelan topik menggunakan LDA yang bersumber dari jurnal, buku, internet dan media lain yang relevan.

#### 3.2. Pengumpulan Data

Data teks diambil dari media sosial yakni Twitter. Pengambilan data dilakukan dengan menentukan kata kunci yang akan dicari terlebih dahulu. Kata kunci tersebut adalah seperti kusta, malaria dan demam berdarah. Pengambilan data dilakukan selama satu bulan yakni di bulan januari 2021 dan diperoleh sebanyak 2737 data.

# 3.3. Praproses Data

Tahap praproses digunakan untuk mempersiapkan data yang akan dianalisis contohnya seperti pembersihan karakter, simbol, teks *noise*, teks yang tidak konsisten dan lain sebagainya. Ada beberapa tahap praproses yang digunakan dalam penelitian ini yaitu:

#### 3.3.1. Proses Tokenization

Data teks yang diperoleh akan dirubah menjadi bagian-bagian yang lebih kecil terlebih dahulu. Pada proses ini dilakukan pemecahan kalimat menjadi tiap-tiap kata.

Proses pemisahan kalimat dapat menggunakan penanda spasi atau tanda baca. Kata-kata yang terbentuk ini disebut dengan token. Kemudian hasil dari proses *tokenization* akan dilakukan ke tahap praproses selanjutnya.

## 3.3.2. Proses Normalization

Proses ini bertujuan untuk mengubah semua data teks menjadi format yang sama. Jadi data teks yang mengandung tanda baca akan dihilangkan, penghapusan imbuhan kata dan merubah huruf besar menjadi huruf kecil.

## 3.3.3. Proses Noise Removal

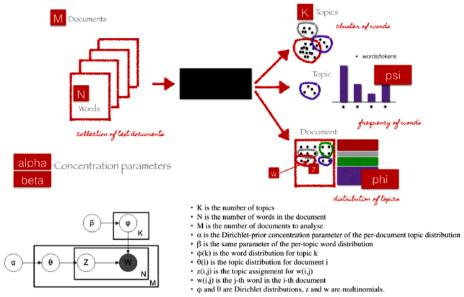
Perolehan data terkadang terdapat teks yang tidak bisa dibaca dan data teks yang mempunyai simbol, tanggal atau waktu, metadata HTML, XML, apapun yang bukan huruf atau sesuatu yang tidak memiliki arti terhadap penelitian maka akan melalui tahap penghapusan agar menjadi sekumpulan kata yang utuh.

# 3.4. Pemodelan Topik

Dasar pemodelan topik yaitu topik yang terbentuk dari kata-kata tertentu dari suatu dokumen (David M. Blei, Andrew Y. Ng, 2003). Pemodelan topik merupakan teknik yang digunakan dalam pendekatan *text mining* dalam menemukan data teks yang tersembunyi dan hubungan antar teks yang saling berkaitan dari suatu korpus (Jelodar & Wang, 2018). Menurut David M. Blei, LDA merupakan model statistik kumpulan dokumen yang mencoba menangkap banyak topik yang saling berkaitan dari dokumen tersebut (Blei, 2012). Dari definisi tersebut maka tujuan utama dari pemodelan topik adalah menemukan topik secara garis besar atau menemukan tema dari sekumpulan teks.

Pemodelan topik termasuk dalam *unsupervised machine learning* dimana didalam pemodelan topik tidak ada label atau definisi teks dan penentuan jumlah topik di tentukan di awal sebelum melakukan training topik modeling. Ada beberapa kemampuan dalam topik modeling yaitu yang pertama secara otomatis dapat melakukan pemindaian dari sekumpulan dokumen, menemukan pola dan topik tersembunyi. Kemudian secara otomatis dapat mengelompokkan kata-kata yang memiliki hubungan satu sama lain dalam satu topik. Terakhir mampu mengasosiasikan token atau kata dan dokumen kedalam topik tersebut.

Salah satu algoritma pemodelan topik adalah *Latent Dirichlet Allocation* (LDA). Ilustrasi algoritma LDA dapat dilihat pada Gambar 2. Secara sederhana algoritma ini berbasis distribusi kata-kata yang ada dalam suatu dokumen. LDA merupakan algoritma yang didasarkan pada *generatif probabilstic model* dimana memiliki dua asumsi yakni setiap topik merupakan campuran dari sekumpulan kata-kata atau sekumpulan token. Dalam hal ini topik direpresentasikan sebagai suatu distribusi kata. Kemudian yang kedua setiap dokumen merupakan campuran dari berbagai macam topik sehingga di dalam satu dokumen akan ada berbagai topik dengan masing-masing probabilitas untuk masing-masing topik tersebut.



Gambar 2 Latent Dirichlet Allocation LDA

Beberapa tahap yang harus dilalui dalam algoritma LDA yakni *initialize parameters*, random topic assignment, resample topic for each word, get result dan model evaluation. Pertama initialize parameter yaitu mendefinisikan semua parameter yang dibutuhkan seperti menentukan jumlah dokumen, topik, iterasi, banyak kata, koefisien LDA (alpha dan beta). Kemudian tahap *random topic assignment*, menandai sebuah kata dengan topik yang telah ditentukan dan dilakukan dengan semi random distribution berdasarkan dirichlet distribution. Selanjutnya resample topic for each word, pada tahap ini dilakukan proses iterasi dimana akan menghasilkan parameter yang dapat menentukan distribusi dari jumlah topik dalam dokumen dan juga distribusi kata dari topik. Untuk setiap kata dalam dokumen ini akan ditetapkan topik secara acak. Dalam tahap ini akan mengevaluasi resample topik, pertama kesesuaian suatu kata di seluruh topik yang ada, apakah keberadaan kata x sesuai dalam topik a, kemudian yang kedua kesesuaian topiktopik dalam dokumen. Dua hal inilah yang akan di evaluasi dalam iterasi. Setelah itu akan diperoleh hasil pemodelan topik dan dari hasil yang diperoleh dapat dilakukan evaluasi terhadap model yang didapatkan.

# 3.5. Analisis dan Evaluasi

Pada tahap ini dilakukan analisis kualitatif dengan memperhatikan *term-term* hasil pengelompokan topik yang sudah terbentuk. Hasil dari pengelompokan topik tersebut akan dianalisis agar dapat diketahui makna terhadap data. Pada tahap evaluasi akan dilakukan penilaian terhadap penelitian yang telah dilakukan yakni seperti kemudahan, kesesuaian dan manfaat. Fokus evaluasi pada penelitian ini adalah evaluasi topik terhadap penyakit tropis di Indonesia khususnya penyakit kusta, malaria dan demam berdarah.

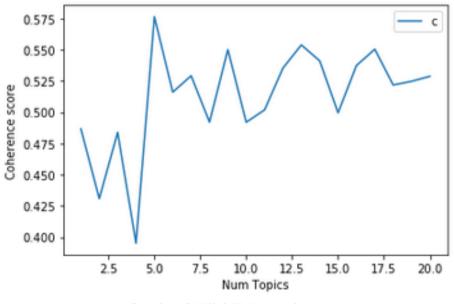
# IV. HASIL DAN PEMBAHASAN

Data yang diperoleh akan melalui tahap praproses guna untuk menstrukturkan, merapikan, membersihkan data yang tidak sempurna, bermasalah, tidak konsisten dan memastikan data siap untuk dianalisis. Pada Tabel 1 merupakan contoh potongan data teks sebelum dan sesudah dilakukan praproses.

Tabel 1 Data Teks Sebelum dan Sesudah Praproses

Sebelum Praproses	Sesudah Praproses
RT @KKMPutrajaya اللَّهُمَّ إِنِّي أَعُوذَ بِكَ مِنَ الْبَرَّصِ، Ya Allah aku Ya Allah aku وَالْجُذُونِ، وَالْجُذَامِ، وَمِنْ سَيِّئِ الْأَسْقَامِ Ya Allah aku berlindung denganMu daripada penyakit sopak, penyakit gila, penyakit kusta dan daripada segala penyakit berbahaya yang lain (Hadith Riwayat Abu Daud) https://t.co/tB9bEY5sVB [RE aryyaff]	penyakit sopak penyakit penyakit kusta penyakit berbahaya
RT Setiap hari yang di blow up media hanya covid. Padahal musim hujan kasus penyakit diare, difteri, malaria, dbd dll naik jumlahnya. https://t.co/ZBBUE6Q4JY [RE Wedhus999]	blow media covid musim hujan penyakit diare difteri malaria dbd
"RT Kenyataan Akhbar KPK 31 Januari 2021 – Sambutan Hari Kusta Sedunia 2021: Bersama HentikanÃ, Kusta	kenyataan akhbar kpk januari sambutan kusta sedunia hentikanaa kusta jabatanpenerangan japenpenang komunikasikita penangjapen

Nilai koherensi digunakan untuk membantu didalam menentukan berapa banyak topik yang tepat dari dokumen yang dimiliki. Semakin tinggi nilai koherensi maka semakin bagus hasil interpretasi topik modeling. Berdasarkan Gambar 3 tiap topik yang dihasilkan mempunyai nilai koherensi yang berbeda.



Gambar 3 Nilai Koherensi

Jumlah nilai koherensi tertinggi berdasarkan data pada Tabel 2 adalah topik ke lima yakni 0.576453. Sehingga hal ini mempunyai arti bahwa jumlah topik yang sesuai dalam dokumen yang dimiliki berjumlah lima. Hasil nilai koherensi pada penelitian ini dapat dilihat pada Tabel 2.

Tabel 2 Nilai Koherensi

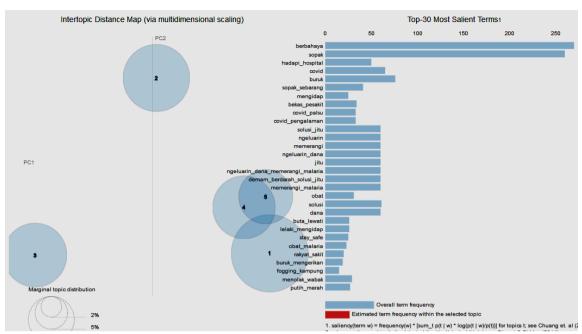
Topik	Nilai koherensi
1	0.486593
2	0.430761
3	0.483805
4	0.395044
5	0.576453
6	0.516031
7	0.529143

Kemudian tiap kata dalam hasil topik modeling akan mempunyai nilai bobot. Nilai bobot tersebut merupakan suatu nilai peluang kemunculan dari tiap kata di dalam topik. Bobot tiap kata dapat dilihat pada Tabel 3.

Tabel 3 Nilai Bobot

Topik	Kata
T1	0.023*"mengidap" + 0.016*"memerangi_malaria" + 0.016*"memerangi" + 0.016*"ngeluarin_dana_memerangi_malaria" + 0.016*"demam_berdarah_solusi_jitu" + 0.016*"ngeluarin" + 0.016*"solusi_jitu" + 0.016*"jitu" + 0.016*"ngeluarin_dana" + 0.016*"solusi"
T2	0.020*"sopak_sebarang" + 0.019*"ngeluarin_dana" + 0.019*"ngeluarin" + 0.019*"solusi_jitu" + 0.019*"solusi" + 0.019*"jitu" + 0.019*"memerangi" + 0.019*"dana" + 0.019*"ngeluarin_dana_memerangi_malaria" + 0.019*"demam_berdarah_solusi_jitu"
Т3	0.037*"buruk" + 0.036*"covid" + 0.031*"sopak" + 0.030*"hadapi_hospital" + 0.020*"bekas_pesakit" + 0.020*"covid_palsu" + 0.020*"covid_pengalaman" + 0.015*"stay_safe" + 0.012*"hospital" + 0.012*"berbahaya"
T4	0.020*"obat" + 0.015*"obat_malaria" + 0.011*"chikungunya" + 0.010*"lakukan_fogging" + 0.010*"dbd" + 0.009*"dengue" + 0.009*"anti" + 0.009*"waspada" + 0.009*"nyamuk" + 0.008*"malarindu"
Т5	0.167*"berbahaya" + 0.134*"sopak" + 0.017*"lelaki_mengidap" + 0.017*"buta_lewati" + 0.014*"rakyat_sakit" + 0.013*"buruk_mengerikan" + 0.011*"sakit" + 0.010*"rumah_sakit" + 0.010*"penderita" + 0.009*"dibangun_sejarah"

Setelah melakukan pemodelan topik LDA, maka hasil pemodelan akan ditampilkan dalam bentuk visualisasi *intertopic distance map*. Dalam visualisasi ini terdapat topik dan kata yang paling banyak muncul. Adapun bentuk visualisasi ini seperti pada Gambar 4.



Gambar 4 Intertopic Distance Map

Pada Gambar 4 menjelaskan 30 kata penting yang muncul di corpus. Terdapat dua sisi dalam visualisasi ini yakni sisi panel kiri dan kanan. Panel sebelah kiri menunjukkan jarak antar topik. Terdapat 3 topik yang saling berdekatan yakni topik 1, topik 4 dan topik 5. Hal ini menunjukkan adanya kesamaan dan kedekatan pembahasan antara ketiga topik tersebut. Sedangkan topik 2 dan topik 3 terlihat jauh dari topik-topik lainnya yang mempunyai arti bahwa topik tersebut tidak ada kedekatan dan kesamaan topik. Sedangkan panel kanan mengenai kata-kata dominan yang sering muncul. Kata-kata dominan tersebut seperti berbahaya, sopak, hadapi hospital, covid, buruk. sopak sebarang, mengidap, bekas pesakit dan lain sebagainya.

Visualisasi selanjutnya dengan menggunakan *wordcloud*. Visualisasi *wordcloud* digunakan untuk mempermudah memperoleh wawasan dari topik-topik yang dihasilkan oleh pemodelan (Naury, Fudholi, & Hidayatullah, 2021). Visualisasi ini berdasarkan hasil pemodelan LDA pada Gambar 4 sehingga terdapat lima visualisasi *wordcloud* yang menunjukkan hasil pemodelan topik. Adapun visualisasi *wordcloud* dapat dilihat pada **Error! Reference source not found.** sampai Gambar 9.



Gambar 5 Wordcloud1



Gambar 6 Wordcloud2



Gambar 7 Wordcloud3



Gambar 8 Wordcloud4



Gambar 9 Wordcloud5

Pada penelitian ini juga membahas makna dari hasil pemodelan topik yang diperoleh. Penafsiran makna dari suatu topik dilihat dari berdasarkan 10 kata yang paling dominan seperti yang ditunjukkan pada Tabel 4.

Tabel 4 Hasil Model Topik

Topik	Terms
Topik #1:	Sopak_sebarang, ngeluarin_dana, ngeluarin, solusi_jitu,
Dana untuk	solusi, memerangi, dana,
memerangi malaria	ngeluarin_dana_memerangi_malaria,
dan demam berdarah	demam_berdarah_solusi_jitu, memerangi_malaria
Topik #2:	Buruk, covid, sopak, hadapi_hospital, bekas_pesakit,
Covid-19	covid_palsu, covid_pengalaman, stay_safe, hospital,
	berbahaya
Topik #3:	Berbahaya, sopak, buta_lewati, lelaki_mengidap,
Buta dan kusta	rakyat_sakit, sakit, rumah_sakit, penderita,
	tangan_kakinya_buntung_matanya, belang
Topik #4:	Obat, obat_malaria, chikungunya, lakukan_fogging, dbd,
Pengobatan dan	dengue, anti, waspada, nyamuk, malarindu
penanggulangan	
Topik #5:	Mengidap, memerangi_malaria, memerangi,
Dana untuk	ngeluarin_dana_memerangi_malaria,
memerangi malaria	demam_berdarah_solusi_jitu, ngeluarin, solusi_jitu, jitu,
dan demam berdarah	ngeluarin_dana, solusi

Topik 1 dan 5 memiliki kemunculan kata yang mirip sehingga makna yang cocok untuk topik tersebut adalah kebutuhan dana yang akan digunakan untuk memerangi

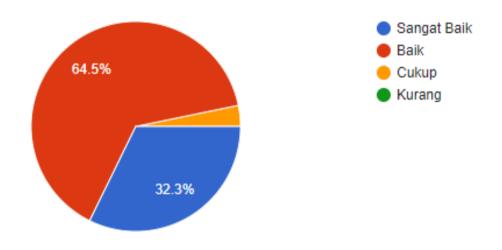
penyakit malaria dan demam berdarah. Hal ini ditandai dengan adanya istilah kata yang muncul seperti dana, memerangi, malaria dan demam berdarah.

Topik 2 menggambarkan kewaspadaan adanya covid-19. Pada topik ini istilah-istilah yang menggambarkan tentang hal tersebut adalah seperti kemunculan kata covid, covid\_palsu, covid\_pengalaman dan *stay safe*. Kemungkinan besar masyarakat Indonesia pada bulan januari 2021 sedang mengalami lonjakan kasus covid-19 dan gejala yang sama dirasakan oleh penyakit demam berdarah yaitu demam tinggi dengan pembuktian terdapat kemunculan kata dbd.

Kemunculan kata seperti buta, belang dan tangan\_kakinya\_buntung\_matanya pada topik 3 menggambarkan penyakit buta dan kusta. Kemunculan kata belang dan buntung memaknai penyakit kusta karena salah satu ciri utama dalam penyakit ini seperti kemunculan lesi pucat, berwarna lebih terang dan menebal serta penderita kusta akut dapat mengakibatkan anggota badan terlepas atau buntung seperti pada bagian ujung jari.

Kemunculnya kata obat dan fogging pada topik 4 sangat tinggi. Sehingga informasi yang paling cocok menggambarkan topik ini adalah pengobatan dan penanggulangan. Hal ini mempunyai arti bahwa masyarakat memperlukan obat untuk penyakit malaria dan demam berdarah serta penanggulangan terhadap penyakit tersebut dengan melakukan penyemprotan (fogging) di daerah masing-masing.

Pengujian penelitian ini menggunakan quisoner dengan responden peneliti, tenaga kesehatan dan akademis. Hasil pengujian pemodelan topik memiliki rata-rata 64.5% mengatakan baik, sangat baik 32.3% dan cukup 3.2% adapun prosentasinya seperti pada Gambar 10**Error! Reference source not found.** 



Gambar 10 Hasil Pengujian

## V. KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Penelitian ini berhasil mengimplementasikan pemodelan topik menggunakan *Latent Dirichlet Allocation* (LDA) terbukti dengan LDA dapat melakukan pemodelan topik terhadap penyakit tropis di Indonesia. Penelitian ini memutuskan banyaknya topik adalah lima, berdasarkan hasil nilai koherensi tertinggi yakni 0.576453. Kemudian hasil pemodelan topik dengan LDA dapat ditampilkan dalam bentuk visualisasi *intertopic distance map* dan *wordcloud*. Berdasarkan penelitian yang telah kami lakukan, hasil pemodelan topik dengan LDA dapat disimpulkan bahwa pembicaraan masyarakat

Indonesia terkait dengan penyakit tropis yakni kusta, malaria dan demam berdarah dengan media Twitter adalah mengenai kebutuhan dana yang digunakan untuk memerangi penyakit malaria dan demam berdarah, covid-19, kebutaan dan kusta, pengobatan dan penanggulangan. Hasil pengujian quisoner dengan responden peneliti, tenaga kesehatan dan akademis, penelitian ini menyatakan 64.5% baik, sangat baik 32.3%.

## 5.2 Saran

Setelah melakukan penelitian pemodelan topik terhadap cuitan penyakit tropis di Indonesia maka penulis menyarankan penelitian selanjutnya agar menggunakan jumlah data yang lebih banyak dan pengembangan metode LDA dengan metode lainnya.

#### **DAFTAR PUSTAKA**

- Alamudi, M. Y. (2018). Pentingnya Eradikasi Penyakit Tropis Di Indonesia. (June).
- Blei, D. M. (2012). Introduction to Probabilistic Topic Models. *Communications of the ACM*, 1–16.
- David M. Blei, Andrew Y. Ng, M. I. J. (2003). Latent Dirichlet Allocation. 3, 993–1022.
- Falah, F., & Syamsidar. (2021). Pengaruh Penerapan Aplikasi Chatbot Sebagai Media Informasi Online Terhadap Kepuasan Pengguna Layanan Kesehatan Primer di Masa Pandemik Covid 19. *Bina Generasi*, (2), 18–23.
- Grover, P., Kumar, A., & Davies, G. (2018). International Journal of Information Management "Technology enabled Health" Insights from twitter analytics with a socio- technical perspective. *International Journal of Information Management*, 43(June), 85–97. https://doi.org/10.1016/j.ijinfomgt.2018.07.003
- Hormansyah, D. S., & Aulia, I. (2018). Penerapan metode tf-idf dan n-gram pada pengembangan aplikasi chatbot berbasis line untuk layanan publik kesehatan di kota malang. *Informatika Polinema*, *5*, 7–11.
- Jelodar, H., & Wang, Y. (2018). Latent Dirichlet Allocation (LDA) and Topic modeling: models, applications, a survey.
- Kang, H. J., & Kim, C. (2019). Analysis of the Trends in Biochemical Research Using Latent Dirichlet Allocation (LDA). 1–14.
- Kesehatan, K., & Indonesia, R. (2019). *Profil Kesehatan Indonesia Tahun 2019*. Indonesia: Kementrian Kesehatan Indonesia.
- Kurniawan, R., & Apriliani, A. (2020). analisis sentimen masyarakat terhadap virus corona berdasarkan opini dari twitter.pdf. *Instek*, 5.
- Mallett, A. J., Quinlan, C., Patel, C., Fowles, L., Crawford, J., Gattas, M., ... Simons, C. (2019). Precision Medicine Diagnostics for Rare Kidney Disease: Twitter as a Tool in Clinical Genomic Translation. *Kidney Medicine*, 1(5), 315–318. https://doi.org/10.1016/j.xkme.2019.06.006
- Naury, C., Fudholi, D. H., & Hidayatullah, A. F. (2021). *Topic Modelling pada Sentimen Terhadap Headline Berita Online Berbahasa Indonesia Menggunakan LDA dan LSTM*. 5, 24–33. https://doi.org/10.30865/mib.v5i1.2556
- Naveen Pemmaraju, Michael A. Thompson, M. Q. (2017). Disease-specific hashtags and the creation of Twitter medical communities in hematology and oncology. *Seminars in Hematology*, 54(4), 189–192. Retrieved from www.sciencedirect.com/science/article/pii/S0037196317301099
- Putra, I. M. K. B. (2017). Analisis Topik Informasi Publik Media Sosial Di Surabaya Menggunakan Pemodelan Latent Dirichlet Allocation (LDA).

- Sarkar, D. (2009). Text Analytics with Python: A Practical Real-World Approach to Gaining Actionable Insights from your Data (Sarkar, ed.). India: Apress.
- Soegijanto, S. (2016). *Kumpulan Makalah Penyakit Tropis dan Infeksi di Indonesia*. Retrieved from https://books.google.co.id/books?hl=id&lr=&id=ZNeFDwAAQBAJ&oi=fnd&pg=P P1&dq=Penyakit+Tropis+Di+Indonesia&ots=dqRj2tLHBp&sig=QUONUNevl6s55 5SMPrax2y1x-kw&redir\_esc=y#v=onepage&q=Penyakit Tropis Di Indonesia&f=false
- Yoga Sahria, & Fudholi, D. H. (2021). Analisis Topik Penelitian Kesehatan di Indonesia Menggunakan Metode Topic Modeling LDA (Latent Dirichlet Allocation). *Resti*, *1*(10), 336–344.