



الجامعة الإسلامية
INDONESIA

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

Dziky Ridhwanullah

17917206

Tesis diajukan sebagai syarat untuk meraih gelar Magister Komputer

Konsentrasi Informatika Medis

Program Studi Informatika Program Magister

Fakultas Teknologi Industri

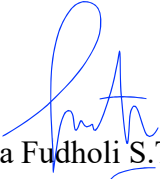
Universitas Islam Indonesia

2022

Lembar Pengesahan Pembimbing

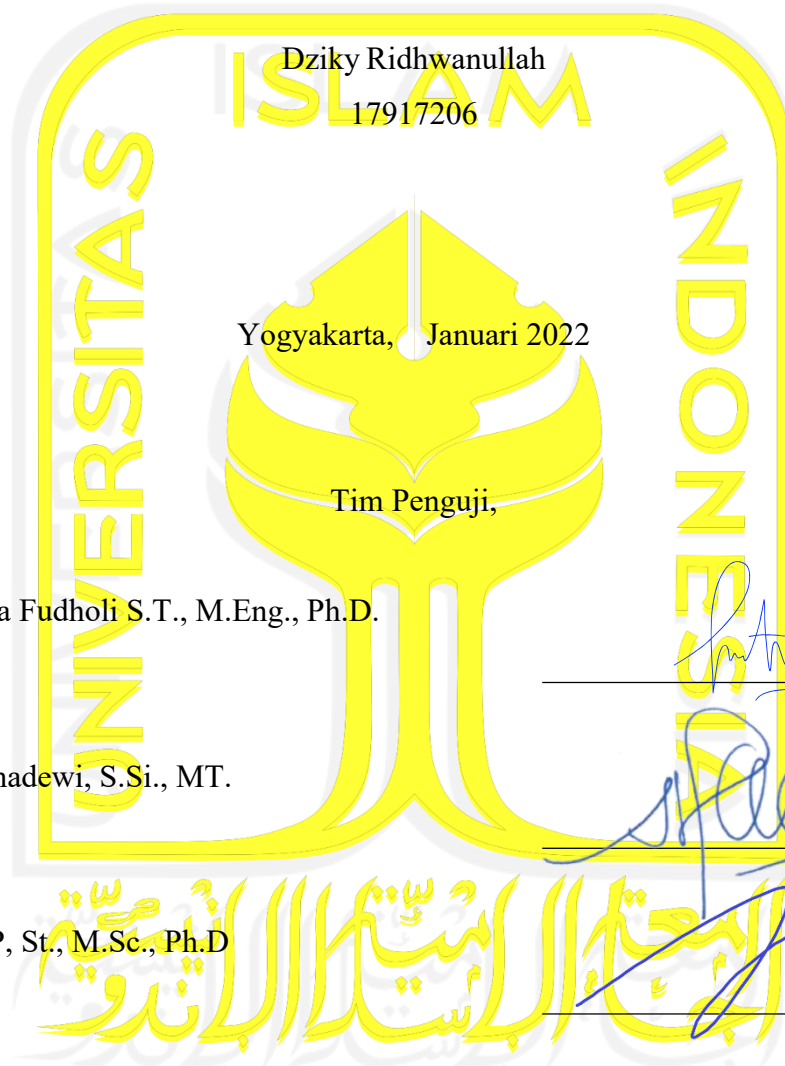
**Pemodelan Topik pada Cuitan tentang Penyakit Tropis di Indonesia dengan
Metode Latent Dirichlet Allocation**




Dhomas Hatta Fudholi S.T., M.Eng., Ph.D.

Lembar Pengesahan Penguji

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



Dziky Ridhwanullah

17917206

Yogyakarta, Januari 2022

Tim Penguji,

Dhomas Hatta Fudholi S.T., M.Eng., Ph.D.

Ketua

Dr. Sri Kusimadewi, S.Si., MT.

Anggota I

Irving Vitra P, St., M.Sc., Ph.D

Anggota II

Mengetahui,

Ketua Program Studi Informatika Program Magister

Universitas Islam Indonesia



Izzati Muhammadiyah, S.T., M.Sc., Ph.D

Abstrak

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

Indonesia memiliki wilayah dan masyarakat yang luas. Sehingga banyak informasi yang muncul melalui media sosial khususnya Twitter. Indonesia secara astronomis termasuk daerah yang memiliki iklim tropis. Sehingga penyakit tropis dapat menyerang Indonesia. Penyakit ini dapat disebabkan oleh bakteri, virus, jamur dan parasit. Pada penelitian ini bertujuan untuk mengetahui topik-topik pembicaraan apa saja yang dibicarakan oleh masyarakat Indonesia terkait dengan penyakit tropis khususnya kusta, malaria dan demam berdarah. Pengambilan data menggunakan teknik *streaming* selama satu bulan yakni pada bulan Januari 2021 dan diperoleh data teks sebanyak 2737. Untuk mengetahui topik-topik pembicaraan dalam data tersebut dapat menggunakan analisis topik modeling. Salah satu metode dalam pemodelan topik adalah dengan menggunakan *Latent Dirichlet Allocation* (LDA). Data tweet penyakit tropis di Indonesia akan dianalisis menggunakan metode ini. Metodenya adalah mencari nilai koherensi terlebih dahulu untuk menentukan jumlah topik yang terbaik. Kemudian dilakukan analisis secara kualitatif berdasarkan hasil tiap topik. Hasil penelitian ini, LDA berhasil memodelkan trend topik pembicaraan masyarakat Indonesia terkait penyakit tropis dan diperoleh sebanyak 5 topik dengan nilai koherensi 0.564049. Berdasarkan hasil pemodelan topik dapat disimpulkan bahwa topik yang dibahas adalah seperti dana yang digunakan untuk memerangi penyakit malaria dan demam berdarah, covid-19, pengobatan dan penanggulangan, perawatan, buta dan kusta.

Kata kunci

Penyakit tropis, pemodelan topik, LDA

Abstract

Modeling Topics on Tweets about Tropical Diseases in Indonesia with the Latent Dirichlet Allocation Method

Indonesia occupies a wide area and society. So that a lot of information emerges through social media, especially Twitter. Indonesia is astronomically among the regions that have a tropical climate. So that tropical diseases can attack Indonesia. The disease can be caused by bacteria, viruses, fungi and parasites. This study intends to determine what topics of conversation are discussed by the Indonesian people related to tropical diseases, especially leprosy, malaria, and dengue fever. The data collecting technique used a streaming technique for one month, specifically in January 2021, and obtained as much as 2737 text data. The analysis of modeling topics can determine the conversation topic in the data. One of the methods in topic modeling is utilizing Latent Dirichlet Allocation (LDA). Tweet data of tropical diseases in Indonesia were analyzed using this method. The method was to find the coherence value first to determine the best number of topics. Then a qualitative analysis was carried out based on the results of each topic. This study result showed that LDA successfully modeled the trend of Indonesian people's conversation topics related to tropical diseases and obtained as many as 5 topics with a coherence value of 0.564049. Based on the results of the topic modeling, it can be concluded that the topics discussed are such as funds used to withstand malaria and dengue fever, COVID-19, treatment and prevention, care, blindness and leprosy.

Keywords

Tropical diseases, topic modeling, LDA

Pernyataan Keaslian Tulisan

Dengan ini saya menyatakan bahwa tesis ini merupakan tulisan asli dari penulis, dan tidak berisi material yang telah diterbitkan sebelumnya atau tulisan dari penulis lain terkecuali referensi atas material tersebut telah disebutkan dalam tesis. Apabila ada kontribusi dari penulis lain dalam tesis ini, maka penulis lain tersebut secara eksplisit telah disebutkan dalam tesis ini.

Dengan ini saya juga menyatakan bahwa segala kontribusi dari pihak lain terhadap tesis ini, termasuk bantuan analisis statistik, desain survei, analisis data, prosedur teknis yang bersifat signifikan, dan segala bentuk aktivitas penelitian yang dipergunakan atau dilaporkan dalam tesis ini telah secara eksplisit disebutkan dalam tesis ini.

Segala bentuk hak cipta yang terdapat dalam material dokumen tesis ini berada dalam kepemilikan pemilik hak cipta masing-masing. Apabila dibutuhkan, penulis juga telah mendapatkan izin dari pemilik hak cipta untuk menggunakan ulang materialnya dalam tesis ini.

ogogyakarta, Januari 2022



Lziky Ridhwanullah

Lziky Ridhwanullah

Daftar Publikasi

Dziky Ridhwanullah, Dthomas Hatta Fudholi. (2022). Pemodelan Topik pada Cuitan tentang Penyakit Tropis di Indonesia dengan Metode Latent Dirichlet Allocation. Jurnal Ilmiah Sinus.

Publikasi yang menjadi bagian dari tesis

Sitasi publikasi 1

Kontributor	Jenis Kontribusi
Dziky Ridhwanullah	Mendesain eksperimen (25%) Menulis <i>paper</i> (30%)
Dthomas Hatta Fudholi	Mendesain eksperimen (35%) Menulis dan mengedit <i>paper</i> (10%)

Halaman Kontribusi

Dalam penyelesaian tesis ini, penulis banyak mendapatkan kontribusi dari berbagai pihak yaitu:

1. Bapak Dhomas Hatta Fudholi S.T., M.Eng., Ph.D. yang berkontribusi sebagai pembimbing tesis secara penuh yang sehingga penulis dapat menyelesaikan studi.
2. Ibu Izzati Muhimmah S.T., M.Sc., Ph.D. sebagai penguji proposal dan progress tesis yang memberikan saran sehingga tesis ini menjadi lebih baik.
3. Bapak Ahmad M. Raf'ie Pratama S.T., M.I.T., Ph.D. sebagai penguji proposal tesis yang telah memberikan arahan terhadap penelitian yang sedang dilakukan
4. Ibu Dr. Sri Kusumadewi S.Si., M.T. sebagai penguji progress dan sidang akhir yang memberikan arahan sehingga penulis dapat lebih mengerti dengan algoritma pada tesis
5. Bapak Irving Vitra P, St.,M.Sc., Ph.D sebagai penguji sidang akhir yang memberikan penguatan terhadap fokus tesis sehingga penulis dapat melakukan beberapa simulasi atau eksperimen dengan algoritma dan memberikan arahan dalam menulis.

Halaman Persembahan

Puji syukur kehadiran Allah SWT yang telah memberikan rahmad, hidayah dan nikmat-Nya sehingga penulis diberikan kemudahan dan kelancaran dalam menyelesaikan penulisan tesis ini dengan baik. Tesis penelitian ini saya persembahkan kepada orang-orang yang telah mendukung dan selalu memberikan motivasi dalam menyelesaikan pendidikan program magister yaitu

1. Kedua orang tua saya yaitu Bapak Suharto dan Ibu Sipah Handayani, adik-adik saya Irfannudien Hidayatullah, Afifah Muftihah, Lulu Nur Azizah dan Qibti Amiroh Rihadatu 'Aisy.
2. Semua keluarga besar Siswomartono yang tidak bisa saya sebutkan satu persatu.
3. Rekan-rekan pascasarjana Magister Informatika konsentrasi Informatika Medis Universitas Islam Indonesia

Kata Pengantar

Syukur alhamdulillah penulis ucapkan dan rahmad, hidayah dan nikmat yang telah diberikan oleh Allah SWT, sehingga studi magister ini bisa diselesaikan dengan baik. Penulis juga bersyukur atas selesainya pengerjaan tesis dengan judul “Pemodelan Topik pada Cuitan tentang Penyakit Tropis di Indonesia dengan Metode Latent Dirichlet Allocation”.

Sholawat dan salam, penulis curahkan untuk Nabi Muhammad SAW. Semoga kita semua mendapatkan syafaat beliau di hari akhirat nantinya, aamiin. Tesis ini dibuat untuk syarat kelulusan program Magister Informatika di Fakultas Teknik Industri di Universitas Islam Indonesia. Penulisan tesis ini berhasil diselesaikan berkat adanya bimbingan dan masukan dari berbagai pihak. Penulis mengucapkan terimakasih kepada pihak-pihak yang telah membantu dalam penyelesaian tesis ini yaitu:

1. Ibu Izzati Muhimmah, ST., M.Sc., Ph.D, selaku Ketua Program Pascasarjana Fakultas Teknologi Industri, Universitas Islam Indonesia, yang selalu memberikan motivasi, semangat dan rasa sayang kepada setiap mahasiswa agar segera menyelesaikan tesis.
2. Bapak Dhomas Hatta Fudholi, S.T., M.Eng., Ph.D. selaku dosen pembimbing yang telah memberikan bimbingan dan saran untuk penulis.
3. Bapak dan ibu dewan penguji sidang seminar proposal, progress dan pendadaran tesis beserta staff tenaga pengajar yang membantu dalam pelaksanaan terkait tesis ini.
4. Teman-teman Magister Teknik Informatika Universitas Islam Indonesia Angkatan 17 yang saling memberikan dorongan dan semangat untuk segera menyelesaikan tesis.
5. Teman peminatan informatika medis yang selalu memberikan masukan.
6. Teman-teman saya yang telah memberikan dorongan dan semangat
7. Pihak-pihak lain yang turut membantu dalam menyelesaikan penelitian ini yang tidak bisa penulis sebut satu persatu.

Penulis berharap semoga Allah SWT membalas semua kebaikan dan kerjasama bapak/ibu. Selanjutnya, penulis juga menyadari bahwa dalam penulisan tesis ini masih jauh dari kesempurnaan, oleh sebab itu penulis berharap saran dan kritik dari pembaca untuk kesempurnaan penulisan berikutnya.

Daftar Isi

Lembar Pengesahan Pembimbing	i
Lembar Pengesahan Penguji.....	ii
Abstrak	iii
Abstract.....	iv
Pernyataan Keaslian Tulisan	v
Daftar Publikasi	vi
Halaman Kontribusi.....	vii
Halaman Persembahan	viii
Kata Pengantar.....	ix
Daftar Isi.....	x
Daftar Tabel.....	xii
Daftar Gambar	xiii
Glosarium	xv
BAB 1 Pendahuluan	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah	3
1.4 Tujuan	3
1.5 Manfaat Penelitian	4
BAB 2 Tinjauan Pustaka	5
2.1 Penyakit	5
2.2 Penyakit Tropis	5
2.2.1 Kusta.....	5
2.2.2 Malaria.....	6
2.2.3 Demam Berdarah.....	7

2.3	Sosial Media	7
2.4	<i>Text Mining</i>	9
2.5	<i>Web Crawling</i>	12
2.6	Pemodelan Topik <i>Latent Dirichlet Allocation (LDA)</i>	14
2.7	Penelitian yang Relevan.....	19
BAB 3 Metodologi		22
3.1	Studi Literatur	22
3.2	Pengumpulan Data.....	22
3.3	Praproses Data	24
3.3.1	Proses <i>Tokenization</i>	24
3.3.2	Proses <i>Normalization</i>	24
3.3.3	Proses <i>Noise Removal</i>	24
3.4	Pemodelan Topik dengan LDA	24
3.5	Analisis dan Evaluasi.....	26
BAB 4 Hasil Dan Pembahasan.....		27
4.1	Praproses Data	27
4.2	Identifikasi frasa	30
4.3	Perhitungan bobot kata	31
4.4	Nilai koherensi	33
4.4.1	Pemodelan Topik dengan LDA	35
4.5	Visualisasi	36
4.6	Hasil dan Analisa	38
4.7	Pengujian	41
BAB 5 Kesimpulan.....		45
5.1	Kesimpulan	45
5.2	Saran	45
Daftar Pustaka		46

Daftar Tabel

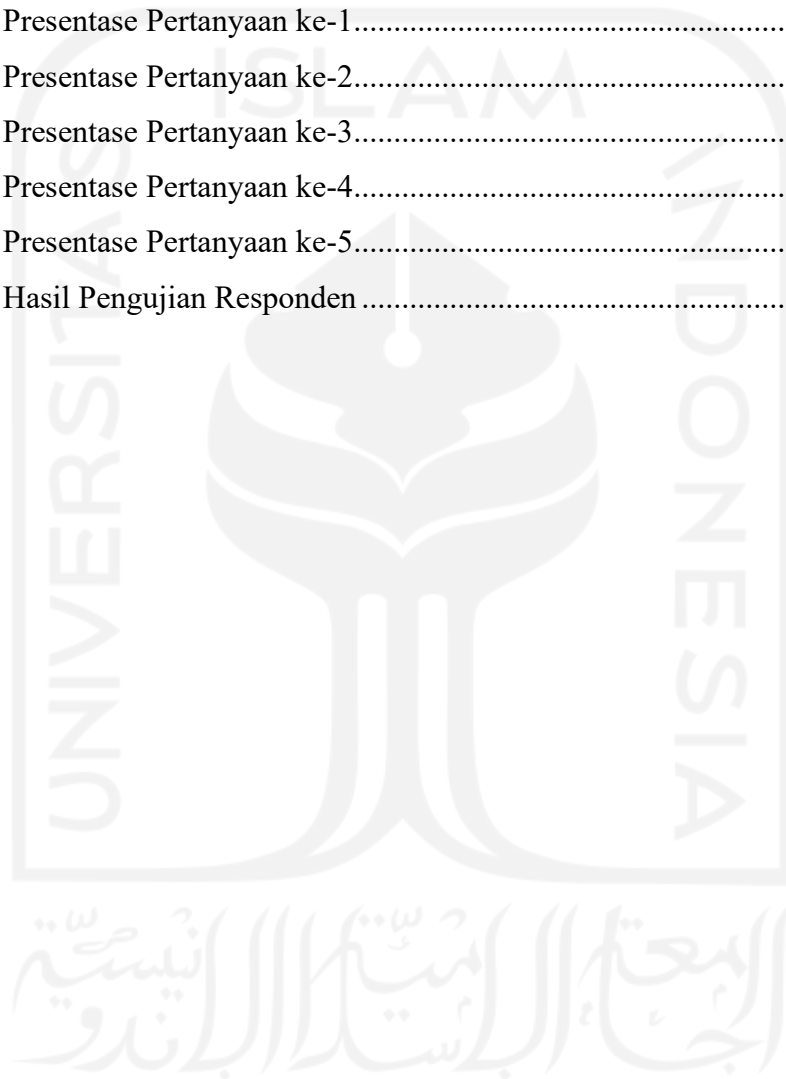
Tabel 2.1 Penelitian Terdahulu.....	19
Tabel 3.1 Dataset Penyakit Tropis.....	23
Tabel 4.2 Perbandingan Data Sebelum dan Sesudah Praproses.....	29
Tabel 4.3 Nilai Koherensi.....	34
Tabel 4.4 Hasil Pemodelan Topik.....	36
Tabel 4.5 Analisa Hasil Pemodelan Topik.....	38
Tabel 4.6 Hasil Pengujian Kuesoner.....	41



Daftar Gambar

Gambar 2.1 Penderita Kusta.....	5
Gambar 2.2 Nyamuk Anopheles	6
Gambar 2.3 Nyamuk Aedes.....	7
Gambar 2.4 Daftar Pengguna Platform Media Sosial	8
Gambar 2.5 Logo Twitter	9
Gambar 2.6 Struktur Data dan Unstruktur Data	10
Gambar 2.7 Proses <i>Text Mining</i>	11
Gambar 2.8 <i>Web Crawling</i>	13
Gambar 2.9 <i>Web Scraping</i>	14
Gambar 2.10 Ilustrasi Algoritma LDA.....	15
Gambar 2.11 Proses LDA.....	16
Gambar 2.12 LDA model	17
Gambar 3.1 Metodologi Penelitian.....	22
Gambar 3.2 <i>Latent Dirichlet Allocation</i> LDA.....	25
Gambar 4.1 Menghapus Kalimat.....	27
Gambar 4.2 Menghapus Alamat Web	28
Gambar 4.3 Menghapus Tanda Baca.....	28
Gambar 4.4 Menghapus angka	28
Gambar 4.5 Menjadikan Huruf Kecil	28
Gambar 4.6 <i>Case Folding</i>	28
Gambar 4.7 <i>Stopwords</i>	29
Gambar 4.8 Bigram dan Trigram	31
Gambar 4.9 Hasil Bigram dan Trigram	31
Gambar 4.10 <i>dictionary</i>	32
Gambar 4.11 Hasil <i>dictionary</i>	32
Gambar 4.12 <i>Bag of Words</i>	32
Gambar 4.13 TF-IDF	32
Gambar 4.14 Hasil TF-IDF	32
Gambar 4.15 Jumlah Topik Berdasarkan Nilai Koherensi ke 1	33
Gambar 4.16 Jumlah Topik Berdasarkan Nilai Koherensi ke 2	33
Gambar 4.17 Jumlah Topik Berdasarkan Nilai Koherensi ke 3	34

Gambar 4.18 LDA	35
Gambar 4.19 Visualisasi <i>Intertopic Distance Map</i>	37
Gambar 4.20 Wordcloud1	38
Gambar 4.21 Wordcloud2	38
Gambar 4.22 Wordcloud3	38
Gambar 4.23 Wordcloud4	38
Gambar 4.24 Wordcloud5	38
Gambar 4.25 Presentase Pertanyaan ke-1.....	42
Gambar 4.26 Presentase Pertanyaan ke-2.....	42
Gambar 4.27 Presentase Pertanyaan ke-3.....	43
Gambar 4.28 Presentase Pertanyaan ke-4.....	43
Gambar 4.29 Presentase Pertanyaan ke-5.....	43
Gambar 4.30 Hasil Pengujian Responden	44



Glosarium

LDA	- Latent Dirichlet Allocation
TF-IDF	- Term Frequency–Inverse Document Frequency
SNA	- Social Network Analysis
API	- Application Programming Interface
HTML	- HyperText Markup Language
URL	- Uniform Resource Locator
HTTP	- Hypertext Transfer – Transfer Protocol



BAB 1

Pendahuluan

1.1 Latar Belakang

Teknologi informasi merupakan suatu kondisi dimana segala sesuatu yang ada dalam kehidupan ini bisa diketahui dimanapun berada dan kapan pun waktunya. Teknologi informasi selalu mengalami perkembangan dari waktu ke waktu. Masyarakat memiliki banyak keuntungan dengan perkembangan teknologi informasi yang ada, mulai dari melakukan sosialisasi melalui media sosial sampai mencari berbagai informasi yang dibutuhkan. Di dalam teknologi informasi setiap orang dimudahkan dalam bersosialisasi yang lebih efektif dan efisien. Manusia dengan berbagai kesibukan dan aktifitasnya masih tetap bisa dengan mudah bersosialisasi menggunakan teknologi informasi. Teknologi informasi menyediakan berbagai kebutuhan manusia dalam bersosialisasi melalui media sosial yang ada baik dari Instagram, Facebook hingga Twitter. Semua media sosial yang ada umumnya menjadi sarana dalam mencari informasi maupun menerima informasi.

Twitter menjadi salah satu media sosial yang sampai saat ini masih di anggap paling cepat dalam memberikan informasi dan komunikasi yang dibutuhkan penggunanya. Di Indonesia sendiri Twitter masih menjadi salah satu media sosial yang memiliki banyak pengguna. Pada tahun 2021 media sosial Twitter di Indonesia menduduki peringkat ke tiga dari enam belas *platform* media sosial lainnya. Twitter sampai saat ini masih menjadi media sosial yang paling di idolakan dalam memperoleh informasi terkini. Informasi paling populer umumnya bisa dengan cepat di peroleh melalui Twitter, hal ini juga yang menjadikan Twitter sering digunakan sebagai sarana bagi banyak orang dalam berdiskusi terutama isu-isu populer dan terkini. Twitter memberikan berbagai macam sajian informasi bagi para penggunanya, mulai dari informasi ekonomi, politik, pendidikan, budaya, maupun kesehatan.

Kesehatan menjadi salah satu isu yang sering diangkat di dalam Twitter oleh para pengguna. Banyak pengguna Twitter yang berdiskusi dan mencari informasi kesehatan di dalamnya, terlebih penyakit yang sedang meningkat kasusnya. Dalam dunia kesehatan, penyakit di kategorikan dalam beberapa kategori yaitu penyakit pada daerah tropis dan subtropis. Indonesia secara astronomis terletak pada 6° Lintang Utara sampai 11° Lintang Selatan dan 95° Bujur Timur hingga 141° Bujur Timur. Kondisi ini yang menjadikan

Indonesia sebagai daerah yang dikategorikan memiliki iklim tropis. Iklim tropis yang terjadi di Indonesia menjadikan negara ini memiliki banyak kasus kesehatan dengan kategori penyakit tropis.

Penyakit tropis merupakan penyakit infeksi yang dapat terjadi didaerah yang memiliki iklim tropis. Penyakit ini dapat disebabkan oleh bakteri, virus, jamur dan parasit (Alamudi, 2018). Penyakit tropis pada umumnya dapat menular. Penyebaran atau penularan dapat terjadi secara langsung antara satu orang ke orang lainnya. Hal tersebut dapat terjadi melalui droplet atau hewan pembawa penyakit seperti nyamuk dan serangga serta faktor lingkungan yang kurang terjaga kebersihannya (Ruminem, Tandirogang, Bakhtiar, Rahayu, & Kadir, 2020). Contoh penyakit tropis seperti malaria, demam berdarah, diare, tetanus, diphteri, amubiasis, demam chikungunya, ascariasis (Soegijanto, 2016). Mengingat penyakit tropis dapat menular ke siapa saja mulai dari anak-anak sampai dewasa serta kemudahan dalam penularan ataupun penyebarannya maka penyakit ini sangat beresiko terjadinya peningkatan kasus.

Kekhawatiran masyarakat terhadap resiko penyebaran dan penularan berbagai penyakit tropis ini menjadikan banyak pengguna Twitter mencari dan memberikan informasi serta berdiskusi, terlebih di musim atau kondisi yang rentan terhadap peningkatan kasus penyakit tropis. Situasi ini umumnya menjadi gambaran bagi sebagian orang untuk mengetahui *trends* topik pembicaraan masyarakat. *Trends* topik yang umumnya menjadi pembicaraan di dalam masyarakat disebabkan oleh beberapa hal yang menjadi pemicunya, kondisi ini menjadikan suatu *trends* topik bisa dikaji secara ilmiah. Dalam kasus *trends* topik pembicaraan masyarakat terkait penyakit tropis ini bisa dikaji secara ilmiah dengan menggunakan analisis pemodelan topik. 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).

Salah satu metode dalam pemodelan topik adalah menggunakan *Latent Dirichlet Allocation* (LDA). Kelebihan LDA adalah menggunakan model probabilistik dengan topik yang mudah diinterpretasikan. LDA dalam penelitian sebelumnya digunakan untuk menganalisis topik terhadap pesan media sosial (Putra, 2017), kondisi cuaca dan iklim (Hidayatullah, Aditya, Karimah, & Gardini, 2019), *headline* berita *online* (Naury, Fudholi, & Hidayatullah, 2021), jurnal penelitian kesehatan (Yoga Sahria & Fudholi, 2021), biokimia (Kang, Kim, & Kang, 2019) dan lain sebagainya.

Karena 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. Nilai koherensi dapat mengukur seberapa mirip kata-kata satu sama lainnya dan untuk menentukan jumlah topik yang terbaik. Kemudian dilakukan analisis secara kualitatif berdasarkan hasil tiap topiknya. Sehingga harapannya dari hasil analisis dapat digunakan pemerintah dan pelayanan kesehatan memberikan informasi maupun gerakan dalam upaya pencegahan atau penyebaran terjadinya kasus penyakit tersebut.

1.2 Rumusan Masalah

Berdasarkan latar belakang, maka rumusan masalah penelitian seperti berikut:

- a. Bagaimana pemodelan topik dengan metode *Latent Dirichlet Allocation* (LDA) terhadap penyakit tropis?
- b. Bagaimana cara menganalisis tren penyakit tropis menggunakan pemodelan topik *Latent Dirichlet Allocation* (LDA)?
- c. Bagaimana memvisualisasikan hasil pemodelan topik dengan metode *Latent Dirichlet Allocation* (LDA) penyakit tropis?

1.3 Batasan Masalah

Berdasarkan latar belakang dan rumusan masalah agar penelitian dapat sesuai dan fokus maka batasan masalah dalam penelitian ini seperti berikut:

- a. Media sosial yang digunakan adalah Twitter.
- b. Pengambilan data dilakukan selama 1 bulan.
- c. Data yang digunakan adalah dari pengguna yang melakukan tweet (menulis/memposting) maupun re-tweet (membagikan).
- d. Fokus penyakit tropis adalah kusta, malaria dan demam berdarah.
- e. Data tweet belum mempertimbangkan orang yang tidak mengekspresikan apa yang dia rasakan sendiri

1.4 Tujuan

Berdasarkan rumusan masalah maka dapat ditentukan tujuan dari penelitian ini adalah sebagai berikut:

- a. Melakukan analisis topik model kesehatan di Indonesia melalui *platform* media sosial.

- b. Dapat memvisualisasikan hasil pemodelan topik penyakit tropis.

1.5 Manfaat Penelitian

Adapun harapan manfaat yang ingin dicapai adalah sebagai berikut:

- a. Hasil penelitian ini diharapkan dapat meningkatkan pemahaman tentang data media sosial twitter dapat digunakan peneliti dalam memahami cuitan atau opini masyarakat terhadap penyakit.
- b. Hasil penelitian ini juga diharapkan berkontribusi dalam penelitian di bidang kesehatan khususnya mengenai bagaimana memodelkan cuitan masyarakat di twitter ke beberapa topik dengan menggunakan metode *Latent Dirichlet Allocation*.
- c. Manfaat untuk masyarakat sebagai masukan pentingnya menyuarakan keadaan perihal deteksi gejala ataupun suatu kasus penyakit khususnya penyakit tropis melalui sosial media twitter sehingga dari hal tersebut dapat digunakan sebagai analisis para peneliti yang kemudian dapat digunakan sebagai acuan pemerintah maupun tenaga kesehatan memberikan upaya atau gerakan atau tindakan nyata ke masyarakat untuk mencegah dan menanggulangi kasus penyakit dan kematian.

BAB 2

Tinjauan Pustaka

2.1 Penyakit

Sakit merupakan adanya ketidakseimbangan fungsi normal pada tubuh manusia. Ada beberapa kriteria penentu keadaan sakit yakni adanya gejala, persepsi keadaan sakit yang dirasa dan menurunnya kemampuan beraktifitas. Beberapa definisi penyakit menurut para ahli dalam bukunya irwan menjelaskan (Irwan, 2017):

- a. Penyakit merupakan kegagalan dari mekanisme adaptasi organisme untuk bereaksi terhadap rangsangan atau tekanan sehingga menimbulkan gangguan pada fungsi/struktur dari bagian sistem tubuh (*Gold Medical Dictionary*).
- b. Penyakit adalah suatu keadaan dimana proses kehidupan tidak lagi teratur atau terganggu perjalanannya (*Van Dale's Woordenboek der Nederlandse Tel*).
- c. Penyakit tidak hanya berupa kelainan saja yang dilihat dapat dilihat dari luarnya akan tetapi suatu keadaan terganggu dari fungsi-fungsi dalam tubuh (*Arrest Hofte Amsterdam*).

2.2 Penyakit Tropis

2.2.1 Kusta

Kusta merupakan penyakit kronik yang disebabkan oleh kuman *mycobacterium leprae* yang menyerang susunan saraf tepi selanjutnya menyerang kulit, mukosa, saluran pernapasan bagian atas, sistem *retikuloendotelial*, mata, otot, tulang dan testis (Amiruddin, 2001). Menurut riwayat lama, penyakit kusta telah ada 6000 tahun SM di Mesir, 600 tahun SM di India dan 500 tahun SM menyebar di Cina (Amiruddin, 2012). Pada Gambar 2.1 merupakan seseorang yang menderita penyakit kusta.



Gambar 2.1 Penderita Kusta

Gejala penderita kusta akan timbul pada stadium lanjut dan dengan diagnosis secara pemeriksaan fisik. Ada 3 tanda-tanda yang menetapkan diagnosis penyakit kusta yakni kelainan kulit yang mati rasa, penebalan saraf tepi yang disertai dengan gangguan fungsi saraf dan terdapat kuman yang tahan asam dalam kerokan jaringan kulit (Amelia, 2020). Seseorang penderita kusta akan diklasifikasikan guna untuk menemukan jenis lamanya pengobatan, waktu penderita dinyatakan sembuh dan perencanaan logistik. Menurut Duanda (1997), klasifikasi ini terdiri dari Indeterminate (IN), Tuberculoid (T), Borderline (B) dan Lepromatous (L).

2.2.2 Malaria

Malaria dikenal sebagai penyakit parasit tropis dan merupakan salah satu dari tiga pembunuh terbesar penyakit menular dan sampai saat ini malaria menjadi masalah kesehatan di dunia (Duarsa, 2020). Malaria disebabkan oleh parasit protozoa dari genus plasmodium yang dimana dapat ditularkan melalui gigitan nyamuk anopheles betina. Malaria dapat ditemukan di hampir seluruh dunia, termasuk wilayah yang mempunyai iklim tropis maupun subtropis. Biasanya malaria menyerang penduduk yang tinggal di daerah endemis atau orang-orang yang berpergian ke daerah yang memiliki angka penularan tinggi. Pada Gambar 2.2 merupakan nyamuk anopheles sebagai pembawa penyakit malaria. Terdapat 400 spesies nyamuk jenis ini akan tetapi hanya 30 sampai 40 spesies yang menyebarkan malaria. Penyakit ini dapat menyerang semua umur dan gejala biasanya akan muncul setelah 10 hari sampai 4 minggu seperti demam, sakit kepala muntah dan menggigil (Supranelfy & Oktarina, 2021).



Gambar 2.2 Nyamuk Anopheles

2.2.3 Demam Berdarah

Penyakit Demam Berdarah Dengue (DBD) menjadi masalah kesehatan di Indonesia yang dimana banyak dijumpai dan menimbulkan Kejadian Luar Biasa (KLB) karena perjalanan penyakit yang cepat dan menyebabkan kematian. Penyakit ini tergolong penyakit yang menular sehingga jumlah penderita meningkat dan penyebarannya semakin luas. Seluruh wilayah Indonesia mempunyai resiko penyebaran penyakit DBD yang disebabkan virus ataupun nyamuk yang sebagai pembawanya.

Pada Gambar 2.3 merupakan nyamuk aedes (*Ae*) *aegypti*. Ciri-ciri nyamuk ini lebih kecil dibandingkan nyamuk lainnya. Nyamuk ini mempunyai dasar hitam dengan bintik-bintik putih pada bagian badan, kaki dan sayapnya. Berdasarkan taksonominya, nyamuk ini termasuk ke dalam ordo diptera, subordo nematocera, famili culicidae, subfamili culicinae, genus aedes, subgenus stegomyia, spesies aedes aegypti (Yulidar & Dinata, 2016).



Gambar 2.3 Nyamuk Aedes

Nyamuk aedes aegypti jenis jantan menghisap cairan tumbuhan atau sari bunga sebagai keperluan kehidupannya dan sedangkan betina menghisap darah. Nyamuk betina ini lebih menyukai darah manusia dibandingkan darah hewan. Biasanya nyamuk ini mencari mangsa pada pagi hari sekitar jam 09.00 sampai 10.00 dan pada petang jam 16.00 – 17.00.

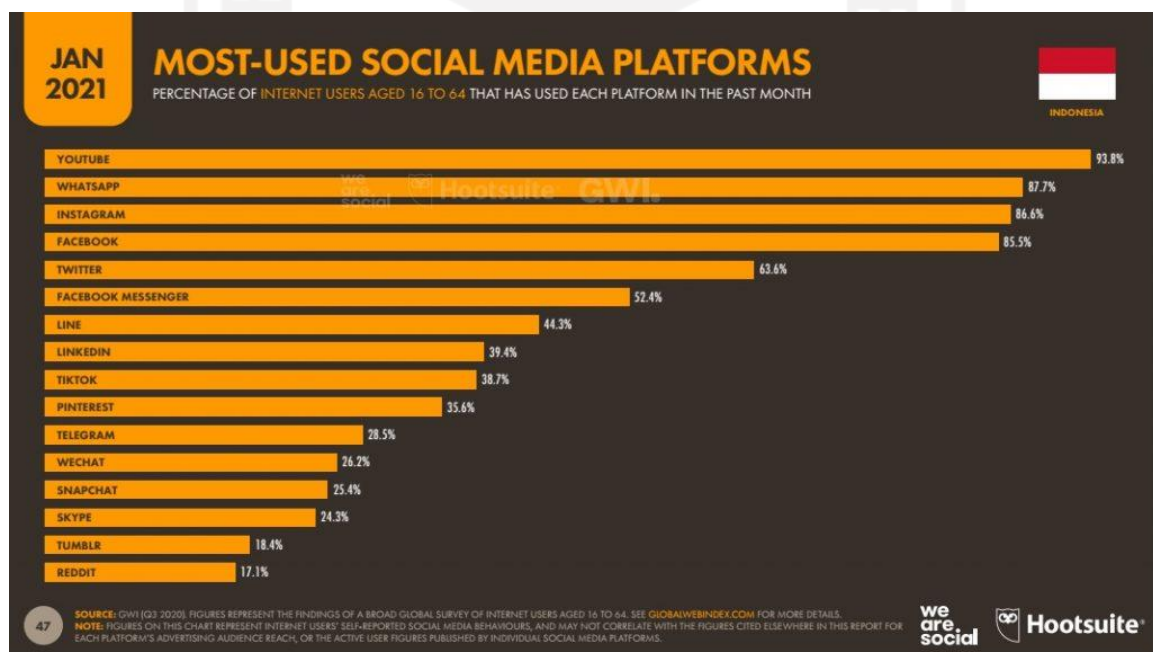
Gejala yang biasa dirasakan oleh penderita apabila terserang penyakit DBD seperti demam tinggi mendadak dengan suhu badan mencapai 38°C sampai 40°C selama 2-7 hari tanpa sebab yang jelas, semangat atau tidak cepat lelah, muncul bintik-bintik merah, kerusakan pembuluh darah dan getah bening, pendarahan di hidung, gusi, dibawah kulit dan muncul memar keunguan (Ratnadewi, Christine, Manurung, Wargasetia, & Dani, 2019).

2.3 Sosial Media

Internet merupakan produk teknologi yang banyak dimanfaatkan oleh masyarakat luas. Pengguna internet di Indonesia pada tahun 2017 mencapai 143,26 juta pengguna. Sebagai

produk teknologi, maka internet dapat memunculkan jenis interaksi sosial baru yang berbeda dengan interaksi sosial sebelumnya (Alyusi, 2016). Pada masa lalu, interaksi terjadi dengan *face to face* maka dengan munculnya teknologi ini, sosial media digunakan masyarakat luas untuk berinteraksi satu dengan lainnya.

Media sosial merupakan destinasi paling banyak yang dikunjungi oleh masyarakat Indonesia. Berdasarkan data dari Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) pada tahun 2020, 89,7% masyarakat menggunakan internet untuk mengunjungi media sosial, komunikasi lewat pesan, mengakses layanan publik, hiburan, layanan perbankan, dan belanja online. Perkembangan media sosial yang semakin pesat dapat memudahkan aktifitas atau pekerjaan manusia. Pada awalnya media sosial hanya dapat digunakan untuk memberitahukan kabar melalui surat, sekarang melalui media sosial dapat memberikan kabar kapan dan dimana saja. Hootsuite (*we are social*) merupakan salah satu situs yang secara berkala menyajikan data dan trend untuk memahami penggunaan internet setiap tahunnya. Berdasarkan survei yang dilakukan oleh situs tersebut di tahun 2021, media sosial yang populer digunakan di Indonesia adalah seperti Youtube, Whatsapp, Instagram, Facebook, Twitter, Facebook Messenger, Line, LinkedIn, Tiktok, Printerest, Telegram, Wechat, Snapchat, Skype, Tumblr dan Reddit seperti pada Gambar 2.4.



Gambar 2.4 Daftar Pengguna Platform Media Sosial

Twitter merupakan salah satu media sosial populer di Indonesia sebesar 63.6%. Pada Gambar 2.5 merupakan gambar yang sangat familiar yakni logo Twitter yang sering

dijumpai. Melalui Twitter pengguna yang tidak terdaftar hanya bisa membaca kicauan sedangkan pengguna terdaftar dapat menulis kicauan (tweet), membagikan kicauan (re-tweet), mengomentari kicauan dan menyukai postingan. Dari fitur tersebut pengguna dapat saling melakukan komunikasi yaitu pengguna dapat berdiskusi, mengutarakan pendapat, memberikan kritikan, menyebarkan informasi dan lain sebagainya. Di awal tahun 2013, Twitter mengangani 500 juta tweet per hari dan 1,6 miliar permintaan pencarian per hari. Sehingga komunikasi yang terus berjalan mempunyai pertumbuhan data yang besar.



Gambar 2.5 Logo Twitter

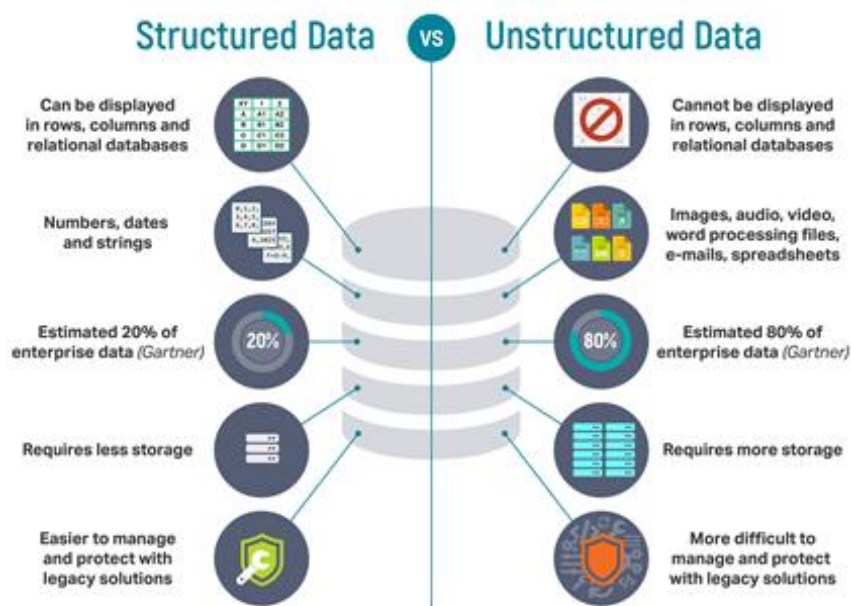
Kompas melangsir bahwa pengguna Twitter pada tahun 2018 mengalami peningkatan yakni pengguna aktif setiap harinya secara global mencapai angka 126 juta (Pratomo, 2019). Oleh sebab itu Twitter semakin memiliki data yang kaya dan *real time* setiap harinya. Developer Twitter melegalkan adanya pemanfaatan data dan menyediakan *Application Programming Interface* (API) guna agar dapat terhubung aplikasi yang dibangun oleh developer lain ke Twitter. Pengguna dapat mengajukan koneksi otorasi dengan Twitter untuk menggunakan Twitter streaming API guna untuk mengotorisasi aplikasi agar dapat mengakses layanan atas nama pengguna tanpa berbagi *password*. Melihat hal ini maka sangat disayangkan jika tidak dimanfaatkan. Sehingga banyak peneliti yang mencoba untuk memanfaatkan data Twitter sebagai bahan penelitian. Contohnya dalam bidang kesehatan seperti mendeteksi prediksi dan memonitoring transisi topik kesehatan dari waktu ke waktu di berbagai daerah (Sidana et al., 2018), pemantauan kesehatan penyakit mental yakni depresi atau bunuh diri (McClellan, Ali, Mutter, Kroutil, & Landwehr, 2017) dan penilaian kesehatan disetiap tweet (Kashyap & Nahapetian, 2014) dan lain sebagainya.

2.4 Text Mining

Data *mining* merupakan ekstraksi informasi atau pola yang penting atau menarik dari data yang ada di *database* yang besar (Siregar & Puspabhuana, 2017). *Text mining* merupakan

salah satu representasi data *mining* karena teks adalah data. Karena cakupan yang sangat luas maka *text mining* dan data *mining* dibedakan. Perbedaan tersebut adalah pola yang digunakan. Pada *text mining* data yang diambil adalah sekumpulan bahasa alami yang memiliki data tidak struktur sedangkan *data mining* adalah kumpulan data yang mempunyai data terstruktur. Dan juga pada penelitian lain mengatakan perbedaan tersebut terdapat pada tugas *mining*-nya seperti klasifikasi, klasterisasi, prediksi dan asosiasi bukan pada karakteristik data yang diolahnya.

Text mining merupakan proses penambangan dan ekstraksi pengetahuan dari data yang berupa teks dan tidak terstruktur dari berbagai sumber seperti dokumen excel, word, pdf, kutipan teks dan lain-lain (Ipmawati, Kusri, & Luthfi, 2017). Data teks dapat ditemui dengan mudah di kehidupan sehari-hari dan data tersebut dapat dianalisis sesuai dengan tujuan penelitian. Contohnya seperti data yang berada pada portal berita media online, sosial media; grup Whatsapp, Line, Telegram, Facebook, Youtube dan Twitter. Data tersebut merupakan data yang bersifat tidak terstruktur seperti data teks, gambar, video, audio dan lain sebagainya. Adapun perbedaan data struktur dengan data tidak struktur seperti pada Gambar 2.6.

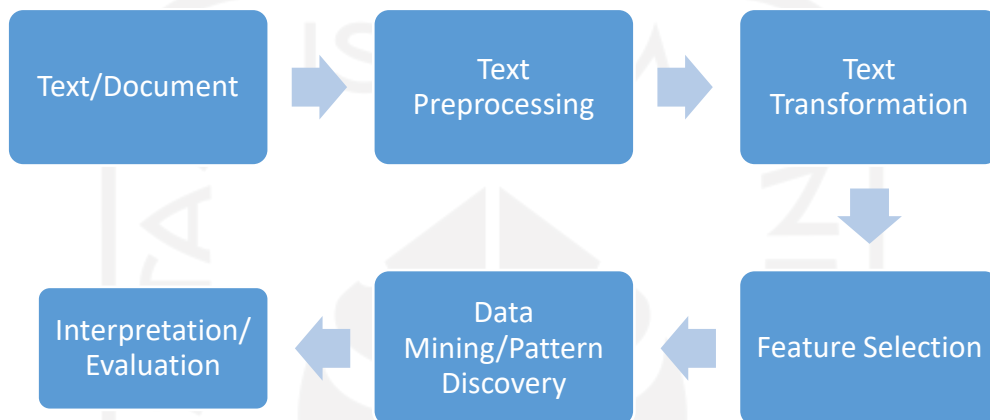


Gambar 2.6 Struktur Data dan Unstruktur Data

Text mining mempunyai banyak fungsi karena cakupannya adalah teks. Oleh karena itu apapun bisa digali berdasarkan informasi yang diperoleh menggunakan *text mining*. Contohnya seperti menganalisis sentimen pada tweet atau re-tweet pengguna Twitter, chat

grup Whatsapp. Kemudian menganalisis *Sosial Network Analysis* (SNA) terhadap kelompok pengguna seperti kelompok pendukung, kelompok kontra dan bahkan bisa mendeteksi kelompok yang dikategorikan sebagai bot.

Ada beberapa tahapan dalam *text mining* seperti pada Gambar 2.7 yakni *text*, *text preprocessing*, *text transformation*, *feature selection*, *data mining/pattern discovery*, *interpretation/evaluation*. Tahap-tahapan tersebut sama dengan tahapan yang ada pada data mining akan tetapi fokus *text mining* lebih banyak pada tahap pre-processing.



Gambar 2.7 Proses *Text Mining*

Proses *text mining* diawali dengan *text* atau *document* yang digunakan sebagai dataset penelitian. Kemudian tahap kedua yaitu *text pre processing* atau praproses digunakan untuk membersihkan dan mempersiapkan dataset yang mengandung teks *noise* dan tidak konsisten. Ada beberapa langkah dalam *pre-processing* yakni proses *tokenization*, *normalization* dan *noise removal*. Tahap *tokenization* semua 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 yang Kemudian tahap *normalization*, tahap 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. Tahap *noise removal*, perolehan data terkadang terdapat teks yang tidak bisa dibaca dan data teks yang mempunyai simbol, tanggal atau waktu, metadata HTML, XML atau apapun yang bukan huruf dan pastinya tidak memiliki arti terhadap penelitian maka hal tersebut akan melalui tahap ini agar menjadi sekumpulan kata yang utuh.

Tahap ketiga dalam *text mining* yakni *text transformation* digunakan untuk mengubah kata ke bentuk dasar dan menghapus kata yang tidak mempunyai arti dalam penelitian. Pendekatan pada tahap ini dapat menggunakan teknik *stemming* dan *stopwords*. Tahap keempat *feature selection*, tahap ini merupakan cara untuk mengoptimalkan kinerja dari classifier. Cara kerja pada tahap ini adalah pengurangan ruang fitur yang besar yakni atribut yang kurang tepat atau tidak relevan akan dihapus dengan menggunakan algoritma *feature selection*. Tahap kelima yakni *process data mining* yang bertujuan untuk memperoleh patterns dari koleksi teks. Pada tahap terakhir adalah *interpretation/evaluation*, tahap ini digunakan untuk *summarization* yakni mengidentifikasi *key content* dari berbagai sumber dimana *key content* dapat merepresentasikan semua teks secara akurat.

2.5 Web Crawling

Sumber data dapat diperoleh dengan menggunakan teknik *web crawling* dan *web scraping*. *Web crawling* merupakan teknik yang digunakan untuk mengambil data dari sebuah website dengan memasukkan alamat *Uniform Resource Locator* (URL). *Web crawling* dimulai dengan daftar URL yang akan dikunjungi, proses ini disebut dengan *seeds*. Ketika *web crawling* mengunjungi URL maka *crawler* akan mengidentifikasi semua *hyperlink* yang ada di halaman dan menambahkannya ke daftar URL yang akan dikunjungi, proses ini disebut dengan *crawl frontier*.

Penerapan teknik *crawling* ini menggunakan *automation program* dan membutuhkan *Application Programming Interface* (API) sebagai jalur komunikasi untuk mendapatkan data. Untuk dapat mengambil data dari suatu website perlu adanya akses *key* dari website tersebut. Akses *key* dapat diperoleh dengan mengajukan registrasi ke developer website. Akses *key* tersebut biasanya terdapat beberapa *unique key* yang akan diberikan oleh *developer* website. Penerapan *crawling* dapat menggunakan beberapa teknik yakni *selective crawling*, *focused crawling*, *distributed crawling*, *web dynamic*, *parallel crawling* dan *popularity*. Langkah-langkah *web crawling* dapat dilihat pada Gambar 2.8.



Gambar 2.8 Web Crawling

Sedangkan *web scraping* merupakan suatu teknik yang digunakan untuk mengambil data dari sebuah website melalui proses ekstraksi informasi menggunakan HTTP atau melalui web browser. Penerapan teknik ini dapat menggunakan cara manual atau juga bisa menggunakan *automation program*. Untuk mendapatkan data harus mengetahui *element HTML* ataupun XML pada sebuah website. Kemudian dari *element-element* tersebut dimasukkan ke dalam program guna untuk mencari data yang sesuai dengan id nama atau nama *class* dari HTML. Penerapan *scraping* dapat menggunakan teknik yakni *copy-paste*, *HTML parsing*, *DOM parsing*, *text patern machine*, *google sheet*, *xpath* dan *vertical aggregation*. Langkah-langkah *web scraping* dapat dilihat pada Gambar 2.9.

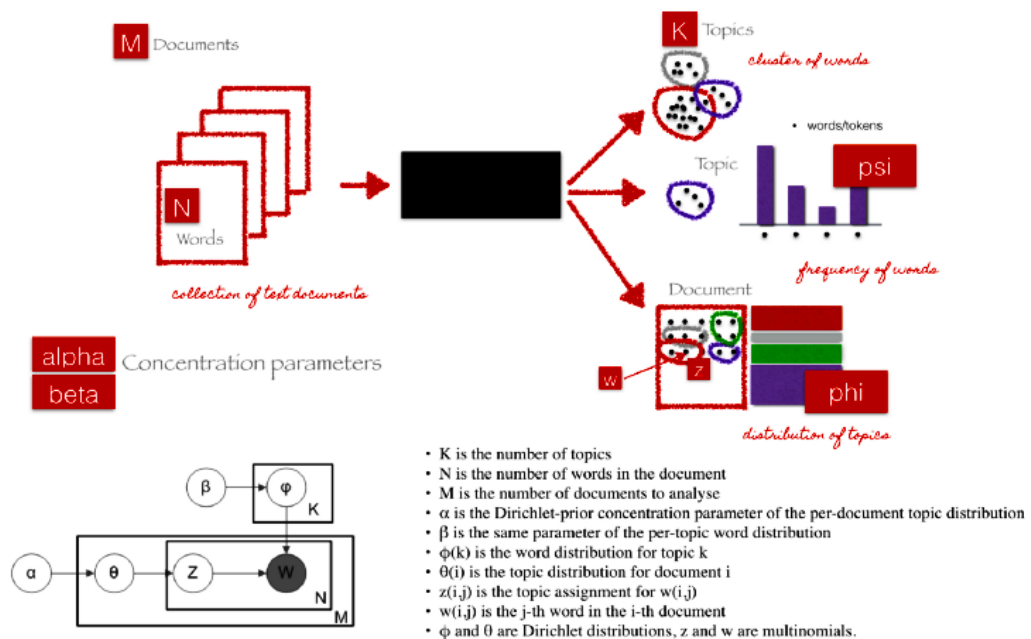


Gambar 2.9 Web Scraping

Ketika menggunakan teknik *web crawling* dalam aktifitas pengambilan data maka aktifitas tersebut secara tidak langsung juga melakukan teknik *web scraping*. Akan tetapi jika menggunakan *web scraping*, maka belum bisa dikatakan melakukan *web crawling*. *Web crawling* biasanya digunakan untuk mengambil data berskala besar dan penerapannya menggunakan API. Sedangkan *web scraping*, biasanya digunakan untuk mengambil data yang berskala kecil atau tidak terlalu besar dan penerapannya menggunakan HTTP.

2.6 Pemodelan Topik *Latent Dirichlet Allocation (LDA)*

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). Dasar pemodelan topik yaitu topik yang terbentuk dari kata-kata tertentu dari suatu dokumen (David M. Blei, Andrew Y. Ng, 2003). Konsep pemodelan topik terdiri dari beberapa entitas yaitu kata, dokumen dan korpora. Kata merupakan unit dasar dari data diskrit dalam dokumen dan kemudian di definisikan sebagai item dari kosa kata yang diberikan suatu indeks di setiap kata dalam dokumen. Kemudian dokumen merupakan susunan N kata-kata dan sebuah korpus adalah kumpulan M dokumen sedangkan korpora merupakan bentuk jamak dari korpus. Salah satu metode pemodelan topik adalah LDA. Pada Gambar 2.10 merupakan ilustrasi pemodelan topik menggunakan algoritma LDA.



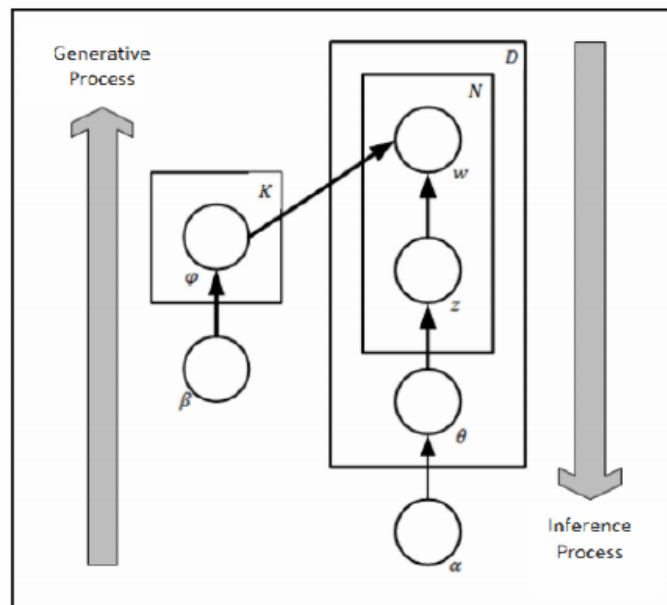
Gambar 2.10 Ilustrasi Algoritma LDA

Pemodelan topik termasuk dalam golongan *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 dan menemukan pola dan juga 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.

Secara sederhana algoritma LDA berbasis distribusi kata-kata yang ada dalam suatu dokumen. LDA akan mencari kata-kata tersebut berasal dari topik yang sama atau tidak. LDA merupakan algoritma yang didasarkan pada *generatif probabilistic* model dimana memiliki dua asumsi yakni masing-masing topik, 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.

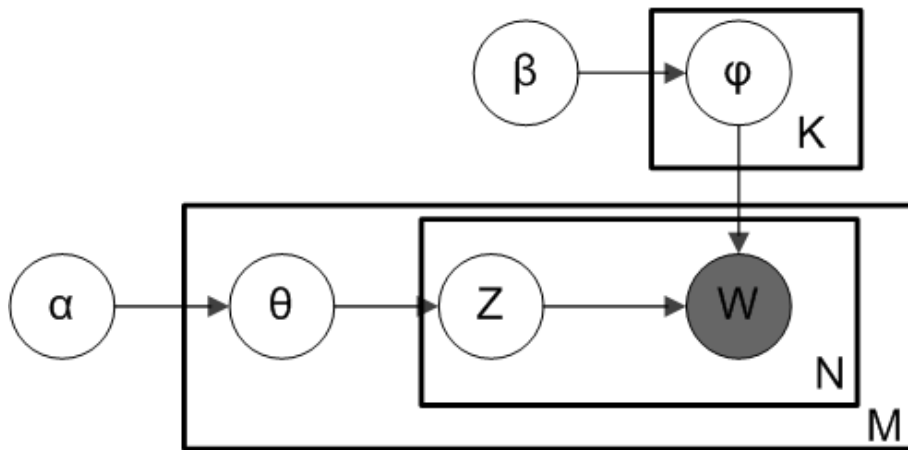
Algoritma LDA terdapat dua proses yaitu LDA sebagai proses *generatif probabilistik* dan LDA sebagai proses inferensi. Pertama LDA yang sebagai proses *generatif probabilistik* akan diimplementasikan saat membuat corpus, seperangkat dokumen, nilai distrbusi kata untuk setiap topik dan proporsi topik setiap dokumen. Kemudian yang kedua, LDA sebagai

proses inferensi diimplementasikan ketika akan mengidentifikasi variabel *latent* termasuk distribusi kata disetiap topik dan proporsi topik dalam dokumen seperti yang digambarkan pada Gambar 2.11.



Gambar 2.11 Proses LDA

Ada beberapa tahap yang harus dilalui dalam algoritma LDA yakni *initialize parameters*, *random topic assignment*, *resample topic for each word*, *get result* dan *model evaluation*. *Initialize parameter* yaitu mendefinisikan semua parameter yang dibutuhkan seperti menentukan jumlah dokumen, topik, iterasi, banyak kata, koefisien LDA (alpha dan beta). Kemudian pada tahap *random topic assignment*, menandai sebuah kata dengan topik yang telah ditentukan dan dilakukan dengan semi random distribution berdasarkan dirichlet distribution. Tahap *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 yakni yang pertama kesesuaian suatu kata di seluruh topik yang ada. Apakah kata x sesuai berada dalam topik a. Kemudian yang kedua, kesesuaian topik-topik 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. Berikut pada Gambar 2.12 merupakan model algoritma LDA.



Gambar 2.12 LDA model

Keterangan:

- α : Dirichlet Distribution
- β : Dirichlet Distribution
- θ : Multinomial distribution
- φ : Multinomial distribution
- Z : Topik
- W : Kata
- N : Banyaknya kata dalam dokumen
- M : Banyaknya dokumen

Dari multinomial distribution dapat membuat topik dan sekumpulan kata. Simbol α merupakan *per-document topic distribution*. α sebagai kontroler distribusi topik setiap dokumen. Semakin tinggi nilai α maka menunjukkan dokumen tersebut memuat topik campuran. Kemudian jika nilai α yang semakin rendah maka menunjukkan tidak banyak topik yang tercampur satu sama lain dalam satu dokumen. β merupakan *per-topic word distribution* dimana menggambarkan distribusi kata di setiap topik. Nilai β yang tinggi menunjukkan bahwa topik x juga memuat kata-kata yang terdapat pada topik lain. Sebaliknya jika nilai β rendah maka sebuah topik akan mengidentifikasi kata-kata yang lebih spesifik. Simbol θ merepresentasikan probabilitas dokumen m . simbol φ merupakan distribusi kata untuk topik tertentu atau merepresentasikan probabilitas topik ke i yang mengandung kata tertentu. Z adalah topik untuk kata ke- n dari suatu dokumen. W adalah spesifik kata dan W adalah kata yang diobservasi.

$$P(W, Z, \theta, \varphi; \alpha, \beta) = \prod_{j=1}^M P(\theta_j; \alpha) \prod_{i=1}^N P(\varphi_i; \beta) \prod_{t=1}^N P(Z_{j,t} | \theta_j) P(W_{j,t} | \varphi_{Z_{j,t}}) \quad (1)$$

Berdasarkan rumus di atas agar mudah dipahami maka dibagi menjadi dua bagian yakni bagian kiri dan kanan. Bagian kiri merupakan probabilitas kemunculan sebuah dokumen dan bagian sebelah kanan akan dibagi menjadi empat sub bagian. sub bagian pertama dan ketiga yang mempunyai tanda kota merah digunakan untuk membantu dalam menemukan topik. Selanjutnya bagian ke dua dan ke empat, akan membantu menemukan kata dari suatu artikel atau dokumen. Bagian *dirichlet distribution* pada bagian *topics* akan dihitung probabilitas dari suatu dokumen apakah memuat topik tertentu dan mengasosiasikan dokumen dengan topik yang sesuai. Misalnya dokumen1 memuat probabilitas 90% topik a dan 10% topik b. kemudian pada bagian *words*, menghitung probabilitas dari suatu topik yang memuat kata tertentu dan mengkaitkan topik dengan kata-kata yang sesuai. Misalnya topik1 terdiri dari kata x sekian persen dan kata xz sekian persen. Kemudian *Multinomial distribution* pada bagian *topics*, menghitung probabilitas dari suatu topik berdasarkan nilai distribusi topik dari dokumen. Kemudian pada bagian terakhir atau *words*, menghitung probabilitas suatu kata untuk diasosiasikan ke dalam topik tertentu.

Representasi pemodelan topik dengan LDA dapat menggunakan visualisasi *interstopik distance map* dan *wordcloud*. Dalam visualisasi *intertopic distance map* dapat menggunakan fitur *PyLDAvis* dimana dapat. Fitur ini dirancang untuk membantu pengguna menafsirkan hasil dari sebaran kata tiap topik. Melalui fitur ini dapat dilihat sebaran kata-kata sesuai jumlah kemunculan kata atau perangkingan dari kemunculan tertinggi hingga terendah. Selain itu dalam fitur ini memvisualisasikan kedekatan topik satu dengan topik-topik lainnya. Dalam visualisasi *wordcloud* akan ditampilkan sebaran kata-kata tiap topiknya dan *wordcloud* digunakan untuk membantu pengguna dalam memperoleh wawasan dari topik-topik yang dihasilkan.

2.7 Penelitian yang Relevan

Berikut pada Tabel 2.1 merupakan beberapa penelitian sebelumnya yang berkaitan dengan penerapan pemodelan topik menggunakan algoritma LDA dan usulan penelitian yang diajukan.

Tabel 2.1 Penelitian Terdahulu

Judul penelitian	Peneliti	Metode	Data	Hasil
Analisis Topik Informasi Publik Media Sosial di Surabaya Menggunakan Pemodelan <i>Latent Dirichlet Allocation</i> (LDA)	(Putra, 2017)	LDA	Twitter dan Facebook	Hasil dalam studi ini menghasilkan 10 <i>passes</i> terbaik berdasarkan nilai <i>perplexity</i> dan kestabilan dapat dicapai pada iterasi kesepuluh setiap distribusi topik yang terpilih secara acak dan terbentuk 4 topik berdasarkan analisis nilai <i>perplexity</i>
Analisis Topik Penelitian Kesehatan di Indonesia Menggunakan Metode Topic Modeling LDA (<i>Latent Dirichlet Allocation</i>)	(Yoga Sahria & Fudholi, 2021)	LDA	Jurnal SINTA	LDA dapat melakukan pemodelan topik terhadap judul penelitian di bidang penelitian kesehatan di Indonesia dan hasil pemodelan topik terbagi menjadi dua topik yaitu topik umum dan topik penyakit.
<i>Analysis of the Trends in Biochemical Research Using Latent Dirichlet Allocation</i> (LDA)	(Kang et al., 2019)	LDA	American Chemical Society	Hasil pada studi ini mengatakan LDA dapat mengidentifikasi lima belas topik utama biokimia selama 20 tahun terakhir dan data penelitian dibagi menjadi empat periode untuk mengulang analisis

Judul penelitian	Peneliti	Metode	Data	Hasil
				pemodelan topik untuk jangka waktu tertentu untuk melihat topik mana yang mengalami penurunan atau peningkatan bobot dari waktu ke waktu.
<i>Topic modeling of weather and climate condition on twitter using latent dirichlet allocation (LDA)</i>	(Hidayatullah et al., 2019)	LDA	Twitter	Hasil pemodelan topik terdapat lima topik penting dari akun Twitter BMKG. Topik yang dibahas oleh akun Twitter BMKG menyebarkan informasi tentang informasi cuaca dan prakiraan cuaca; informasi cuaca terkini di Wilayah Yogyakarta; prakiraan cuaca dan peringatan di Jawa Tengah dan Jawa Barat; informasi gempa; artikel terbaru dan informasi cropping kalender.
<i>Topic Modelling pada Sentimen Terhadap Headline Berita Online Berbahasa Indonesia Menggunakan LDA dan LSTM</i>	(Naury et al., 2021)	LDA dan LSTM	Media masa online	LSTM digunakan untuk memberikan sentimen yang menghasilkan tingkat akurasi 71.13% dan pemodelan topik menggunakan LDA dapat menghasilkan topik-topik yang mudah interpretasikan ke dalam visualisasi.

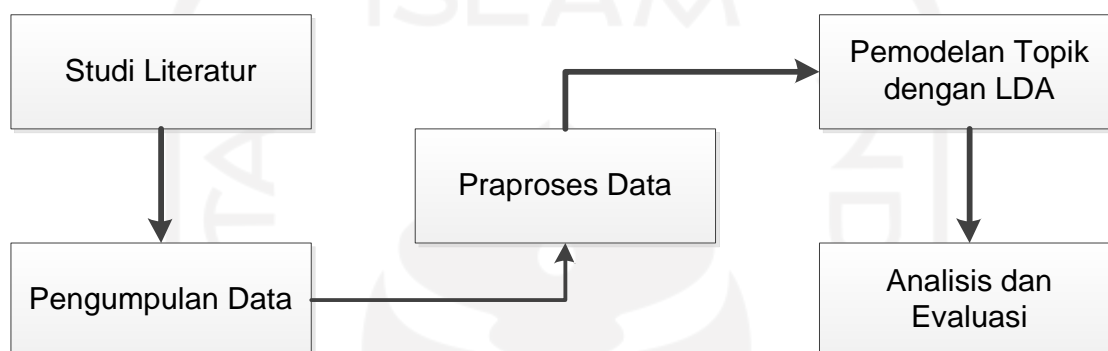
Judul penelitian	Peneliti	Metode	Data	Hasil
Pemodelan Topik pada Cuitan tentang Penyakit Tropis di Indonesia dengan Metode <i>Latent Dirichlet Allocation</i>	Dziky Ridhwanullah Penelitian yang akan diajukan	LDA	Twitter	Hasil pemodelan topik dengan LDA dapat ditampilkan dalam bentuk visualisasi <i>intertopic distance map</i> dan <i>wordcloud</i> .



BAB 3

Metodologi

Bab ini menjelaskan bagaimana penelitian dilakukan sehingga dapat diketahui rincian urutan mengenai tahapan yang dibuat secara sistematis dan dapat digunakan sebagai pedoman dalam menyelesaikan permasalahan dan membuat analisis terhadap hasil penelitian. Metode penelitian yang digunakan seperti pada Gambar 3.1.



Gambar 3.1 Metodologi Penelitian

3.1 Studi Literatur

Dalam penelitian ini studi literatur digunakan untuk mengumpulkan data-data atau sumber yang berhubungan dengan topik penelitian. Studi literatur diperoleh dari berbagai sumber yakni jurnal, buku, internet, dokumentasi dan pustaka baik *online* maupun *offline*. Adapun studi literatur dalam penelitian ini adalah mengenai definisi penyakit, penyakit tropis, media sosial Twitter, *text mining* dan pemodelan topik menggunakan LDA.

3.2 Pengumpulan Data

Target data dalam penelitian ini adalah cuitan masyarakat Indonesia di sosial media khususnya Twitter yang terkait dengan penyakit tropis. Cuitan pengguna yang diambil adalah *tweet*, *re-tweet* dan *reply*. Adapun fokus penyakit tropis seperti penyakit kusta, malaria dan demam berdarah. Kemudian pengambilan data dilakukan dengan menggunakan sistem drone empirit dengan cara mengajukan kata kunci yang akan dicari. Kata kunci tersebut adalah seperti kusta, malaria dan demam berdarah. Drone empirit melakukan teknik pengambilan data dengan *streaming* secara *real-time* dan pengambilan data dilakukan selama satu bulan di bulan januari 2021 dan data yang berhasil didapatkan sebanyak 2737. Hasil dari pengumpulan data akan digunakan sebagai dataset penelitian yang berbentuk

document excel dengan format csv. Berikut potongan hasil pengambilan data yang disajikan pada Tabel 3.1

Tabel 3.1 Dataset Penyakit Tropis

No	Type	Teks	Date
1	rt	RT @KKMPutrajaya، اللّهُمَّ إِنِّي أَعُوذُ بِكَ مِنَ الْبَرَصِ، وَالْجُنُونِ، وَالْجَذَامِ، 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]	2021-01-31 23:32:43
2	rt	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]	2021-01-31 23:29:33
3	rt	RT Kenyataan Akhbar KPK 31 Januari 2021 “ Sambutan Hari Kusta Sedunia 2021: Bersama Hentikan, Kusta #JabatanPenerangan #JapenPenang #KomunikasiKita [RE PenangJapen]	2021-01-31 23:04:27
4	rt	RT @ifansatyaa: Innalilahi, Ya Allah, aku berlindung kepadamu dari penyakit lepra, gila, kusta, dan penyakit- penyakit buruk https://t.co/h [RE ifansatyaa]	2021-01-31 22:55:50
5	rt	RT @KKMPutrajaya، اللّهُمَّ إِنِّي أَعُوذُ بِكَ مِنَ الْبَرَصِ، وَالْجُنُونِ، وَالْجَذَامِ، 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]	2021-01-31 22:37:46
6	rt	RT Temukan Kasus Periksa Kontaknya dan Obati Sampai Tuntas untuk Mencapai Eliminasi Kusta Tahun 2024. Peringatan Hari Kusta Sedunia Tahun 2021 @ganjarpranowo @TajYasinMZ #HariKustaSedunia https://t.co/muuBNeOK6J [RE RSUDKelet]	2021-01-31 22:30:46
7	reply	@deevandra Dulu pernah kena malaria... Di saranin makan sayur yg pahit2, seperti pare n daun pepaya... Alhamdulillah cepat sembuh ya dan jarang kambing lg. $\hat{\sim} \hat{\sim} \hat{\sim} \hat{\sim} \hat{\sim}$	2021-01-31 20:20:52

Pada Tabel 3.1 merupakan potongan hasil pengumpulan data teks dengan isian kolom no, type, teks, date dan lain sebagainya yang berbentuk file berformat csv. Dari kumpulan data ini maka akan dijadikan dataset yang kemudian akan melalui tahap praproses.

3.3 Praproses Data

Setelah melakukan pengambilan data maka data tersebut akan disiapkan terlebih dahulu agar dapat dianalisis. Data tersebut akan melalui tahap praproses. Tahap praproses data digunakan untuk menstrukturkan, merapikan, membersihkan data yang tidak sempurna, bermasalah, tidak konsisten dan memastikan data siap untuk dianalisis. Ada beberapa tahap praproses yang dapat digunakan 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, simbol, frase dan entitas dari sebuah data teks yang nantinya akan dianalisa. Proses pemisahan ini dapat menggunakan penanda spasi atau tanda baca. Kata-kata yang terbentuk dari tahap ini disebut dengan token yang 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 atau apapun yang bukan huruf dan pastinya tidak memiliki arti terhadap penelitian maka hal tersebut akan melalui tahap ini agar menjadi sekumpulan kata yang utuh.

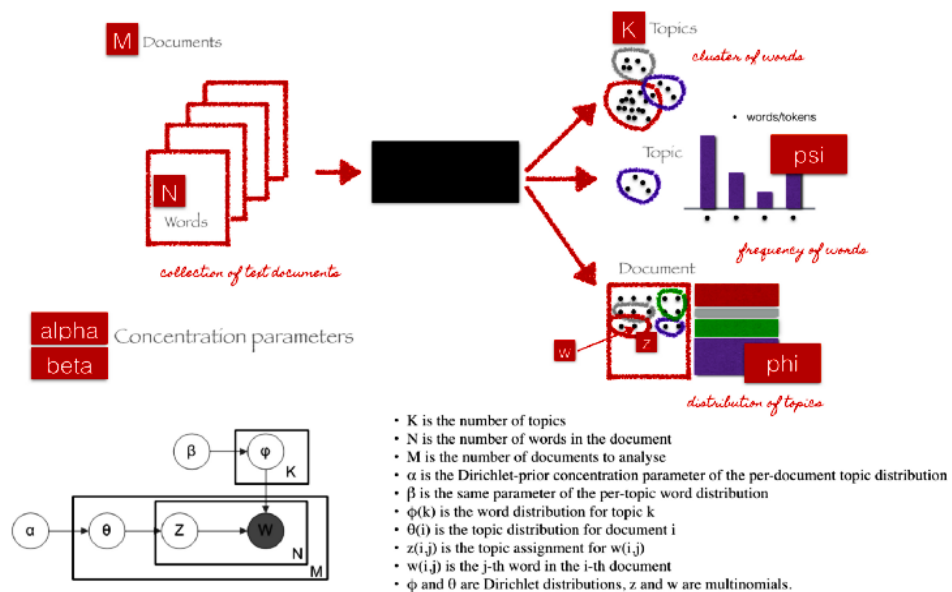
3.4 Pemodelan Topik dengan LDA

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 golongan *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 dan menemukan pola dan juga 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)*. Secara sederhana algoritma ini berbasis distribusi kata-kata yang ada dalam suatu dokumen. LDA akan mencari kata-kata tersebut berasal dari topik yang sama atau tidak. LDA merupakan algoritma yang didasarkan pada *generatif probabilistic* model dimana memiliki dua asumsi yakni masing-masing topik, 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. Pada Gambar 3.2 merupakan ilustrasi algoritma LDA.



Gambar 3.2 *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 topik-topik 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 dan yang kemudian hasil dari analisis dapat digunakan oleh pemerintah, pelayanan kesehatan dan masyarakat.

Pada tahap evaluasi akan dilakukan penilaian terhadap penelitian yang telah dilakukan. Pada tahap ini menggunakan quisoner dengan responden peneliti, tenaga kesehatan dan akademis. Fokus evaluasi pada penelitian ini adalah evaluasi topik terhadap penyakit tropis di Indonesia khususnya penyakit kusta, malaria dan demam berdarah.

BAB 4

Hasil Dan Pembahasan

Pada Bab ini akan dibahas mengenai implementasi metode yang dibahas pada Bab sebelumnya. Adapun bahasan pada Bab ini seperti praproses data, identifikasi frasa, perhitungan bobot, LDA, visualisasi, hasil dan analisa.

4.1 Praproses Data

Data yang diperoleh maka akan melalui tahap praproses data guna untuk menstrukturkan, merapikan, membersihkan data yang tidak sempurna, bermasalah, tidak konsisten dan memastikan data siap untuk dianalisis. Praproses merupakan langkah yang sangat penting karena berpengaruh dengan hasil pemodelan topik. Langkah-langkah yang dilakukan dalam tahap ini dapat meningkatkan kualitas pemodelan topik. Praproses data dilakukan dengan bahasa python versi 3.7. Adapun dalam penelitian ini dilakukan praproses data dengan beberapa tahap yaitu:

a. Menghapus kalimat

Data teks atau dataset yang diperoleh biasanya juga terdapat teks yang mempunyai beberapa kalimat saja misal satu sampai lima kalimat. Kemungkinan besar data tersebut tidak mempunyai arti terhadap penelitian. Oleh karena itu akan dilakukan proses penghapusan terhadap kalimat tersebut. Dalam penelitian ini data teks yang hanya memiliki satu kalimat saja akan dihapus. Adapun implementasi kode program seperti yang ditunjukkan pada Gambar 4.1.

```
#remove sentence which contains only one word
def removeSentence(str):
    word = str.split()
    wordCount = len(word)
    if(wordCount<=1):
        str = ''
    return str
```

Gambar 4.1 Menghapus Kalimat

b. Menghapus alamat web

Tahap ini dilakukan guna untuk menghapus alamat website yang berada pada data teks. Berikut potongan implementasi kode program yang ditunjukkan pada Gambar 4.2.

```
str = re.sub(r'(?i)\b(?:https?://|www\d{0,3}[.]|[a-z0-9.\-]+[.][a-z]{2,4}/)
```

Gambar 4.2 Menghapus Alamat Web

c. Menghapus tanda baca

Tahap ini dilakukan untuk menghapus tanda baca seperti ?!.,:” dan lain sebagainya. Adapun potongan kode program seperti pada Gambar 4.3.

```
str = re.sub(r'^\w|_|_', ' ', str)
```

Gambar 4.3 Menghapus Tanda Baca

d. Menghapus angka dalam teks dan angka

Tahap ini dilakukan untuk menghapus data angka didalam teks misalnya seperti “aku123” dan menghapus data angka seperti 1,2,3,4,5 dan seterusnya. Berikut implementasi kode program yang ditunjukkan pada Gambar 4.4.

```
#remove digit from string  
str = re.sub("\S*\d\S*", "", str).strip()  
#remove digit or numbers  
str = re.sub(r"\b\d+\b", " ", str)
```

Gambar 4.4 Menghapus angka

e. Menjadikan huruf kecil

Kemudian pada tahap ini semua data teks akan dijadikan huruf kecil yang kemudian hasil tahap ini akan dilanjutkan ke tahap berikutnya. Adapun implementasi kode program seperti pada Gambar 4.5.

```
str = str.lower()
```

Gambar 4.5 Menjadikan Huruf Kecil

f. Case folding

Setelah itu tahap ini dilakukan penyamaan bentuk huruf dari keseluruhan teks menjadi huruf kecil semua. Berikut potongan kode program yang ditunjukkan pada Gambar 4.6.

```
str = str.lower()
```

Gambar 4.6 Case Folding

g. Menghapus *stopwords*

Pada tahap ini dilakukan penghapusan kata-kata yang dianggap tidak penting dan tidak mempunyai pengaruh yang signifikan terhadap proses pemodelan topik pada penelitian ini. Kata-kata tersebut dapat disesuaikan menurut pilihan bahasa. Pada penelitian ini pemilihan *stopwords* menggunakan pilihan bahasa Indonesia. Setelah memperhatikan hasil pemodelan topik, maka dapat dievaluasi kata-kata yang tidak penting terhadap penelitian yang dilakukan, sehingga selanjutnya dilakukan penambahan kata-kata tersebut dan ditambahkan ke dalam file *stopwords* yang ada di folder *corpora*. Akses penempatan penambahan kata dapat dilakukan *drive* C:\users\..\AppData\Roaming\nltk_data\corpora\stopwords. Berikut potongan kode program yang ditunjukkan pada Gambar 4.7.

```
def removeStopword(str):
    stop_words = set(stopwords.words('indonesian'))
    word_tokens = word_tokenize(str)
    filtered_sentence = [w for w in word_tokens if not w in stop_words]
    return ' '.join(filtered_sentence)
```

Gambar 4.7 *Stopwords*

Berikut merupakan perbandingan data teks sebelum dan sesudah melalui tahap praproses yang dapat dilihat pada Tabel 4.1.

Tabel 4.1 Perbandingan Data Sebelum dan Sesudah Praproses

Data Teks Sebelum Praproses	Data Teks Setelah Praproses
RT @KKMPutrajaya اللَّهُمَّ إِنِّي أَعُوذُ بِكَ مِنَ الْبَرَصِّ، وَالْجُنُونِ، وَالْجُدَامِ، 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

Data Teks Sebelum Praproses	Data Teks Setelah Praproses
RT Kenyataan Akhbar KPK 31 Januari 2021 “ Sambutan Hari Kusta Sedunia 2021: Bersama Hentikan, Kusta #JabatanPenerangan #JapenPenang #KomunikasiKita [RE PenangJapen]	kenyataan akhbar kpk januari sambutan kusta sedunia hentikanaa kusta jabatanpenerangan japenpenang komunikasikita penangjapen
RT @ifansatyaa: Innalilahi, Ya Allah, aku berlindung kepadamu dari penyakit lepra, gila, kusta, dan penyakit-penyakit buruk https://t.co/h [RE ifansatyaa]	ifansatyaa innalilahi kepadamu penyakit lepra kusta penyakit penyakit buruk ifansatyaa
RT @KKMPutrajaya اللَّهُمَّ إِنِّي أَعُوذُ بِكَ مِنَ الْبَرَصِ، وَالْجُنُونِ، وَالْجُدَامِ، 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 Temukan Kasus Periksa Kontaknya dan Obati Sampai Tuntas untuk Mencapai Eliminasi Kusta Tahun 2024. Peringatan Hari Kusta Sedunia Tahun 2021 @ganjarpranowo @TajYasinMZ #HariKustaSedunia https://t.co/muuBNeOK6J [RE RSUDKelet]	temukan periksa kontaknya obati tuntas mencapai eliminasi kusta peringatan kusta sedunia ganjarpranowo tajyasinmz harikustasedunia rsudkelet
@deevandra Dulu pernah kena malaria... Di saranin makan sayur yg pahit2, seperti pare n daun pepaya... Alhamdulillah cepet sembuh ya dan jarang kambing lg. [RE deevandra]	deevandra malaria saranin makan sayur pare daun pepaya alhamdulillah cepet sembuh jarang kambing

4.2 Identifikasi frasa

Setelah keseluruhan data melalui tahap praproses maka kemudian dilakukan pembuatan model bigram dan trigram. Bigram merupakan dua kata yang sering muncul dalam suatu dokumen sedangkan trigram merupakan tiga kata yang sering muncul. Tujuan tahap identifikasi frasa adalah untuk mengidentifikasi frasa dengan melihat kemunculan dua atau

tiga kata yang sering muncul bersamaan Implementasi kode program pada tahap ini seperti pada Gambar 4.8.

Pada tahap ini akan dilakukan pencarian dan penggabungan dari semua kata berdasarkan kemunculan dalam bentuk bigram dan trigram. Penggabungan dengan menggunakan tanda *underscore* yakni seperti kata demam_berdarah. Setiap bigram dan trigram merupakan sebuah frasa. Kemudian frasa tersebut akan dimasukkan ke dalam satu kalimat dan hasil pada tahap ini dapat dilihat pada Gambar 4.9.

```
import gensim
#Create Bigram & Trigram Models
from gensim.models import Phrases
# Add bigrams and trigrams to docs, minimum count 10
bigram = Phrases(text_list, min_count=10)
trigram = Phrases(bigram[text_list])
for idx in range(len(text_list)):
    for token in bigram[text_list[idx]]:
        if '_' in token:
            # Token is a bigram, add to document.
            text_list[idx].append(token)
    for token in trigram[text_list[idx]]:
        if '_' in token:
            # Token is a bigram, add to document.
            text_list[idx].append(token)
```

Gambar 4.8 Bigram dan Trigram

```
'demam_berdarah', 'demam_berdarah', 'demam_berdarah'], ['penyakit'
. 'penyakit', 'berbahaya'], ['penyakit', 'sopak', 'penyakit', 'pen
. ['anak', 'panu', 'kunjung', 'sembuh', 'waspada', 'penyakit', 'ku
```

Gambar 4.9 Hasil Bigram dan Trigram

4.3 Perhitungan bobot kata

Pada tahap ini akan dilakukan pembuatan *dictionary* yang dimana akan merepresentasikan isi dari dokumen. Pembuatan *dictionary* dapat menggunakan library dari gensim yakni modul *corpora.Dictionary()*. Parameter pada modul ini berisi data teks array dua dimensi yakni data cuitan penyakit tropis malaria, demam berdarah dan kusta. Implementasi kode program membuat *dictionary* dapat dilihat pada Gambar 4.10 dan hasil pembuatan *dictionary* dapat dilihat pada Gambar 4.11.


```

from gensim import corpora, models
# Create a dictionary representation of the documents.
dictionary = corpora.Dictionary(text_list)
dictionary.filter_extremes(no_below=5, no_above=0.2)
#no_below (int, optional) - Keep tokens which are contained
#no_above (float, optional) - Keep tokens which are contained
print(dictionary)

```

Gambar 4.10 *dictionary*

```

Dictionary(1018 unique tokens: ['berbahaya', 'sopak', 'blow',
'blow_media', 'covid']...)

```

Gambar 4.11 Hasil *dictionary*

Kemudian langkah selanjutnya adalah membuat *bag of words*. *Bag of words* merupakan metode yang digunakan untuk mengubah teks ke dalam bentuk vektor agar dapat dikenali oleh komputer. Pembuatan *bag of words* dapat menggunakan fungsi dari *doc2bow* dimana tujuannya untuk mengkonversi dokumen menjadi format matriks, menghitung jumlah kemunculan setiap kata yang berbeda, mengkonversi kata ke id kata bilangan dan mengembalikan hasilnya ke bentuk vektor. Adapun potongan penggunaan kode program *bag of word* dapat dilihat pada Gambar 4.12.

```

doc_term_matrix = [dictionary.doc2bow(doc) for doc in text_list]
print(len(doc_term_matrix))
print(doc_term_matrix[100])

```

Gambar 4.12 *Bag of Words*

Setelah membuat *bag of words*, akan dilakukan perhitungan bobot untuk setiap kata. Pemberian bobot tersebut dapat menggunakan metode TF-IDF. *Term Frequency–Inverse Document Frequency* (TF-IDF) merupakan ukuran statistik yang menggambarkan kata (*term*) dalam dokumen dan korpus. Adapun potongan penggunaan kode program TF-IDF dapat dilihat pada Gambar 4.13 dan hasil proses ini dapat dilihat pada Gambar 4.14.

```

tfidf = models.TfidfModel(doc_term_matrix) #build TF-IDF model
corpus tfidf = tfidf[doc term matrix]

```

Gambar 4.13 TF-IDF

```

2737
[(143, 1), (144, 1), (145, 1), (146, 1), (147, 2), (148, 1),
(149, 1), (150, 1), (151, 2), (152, 1), (153, 1), (154, 1)]

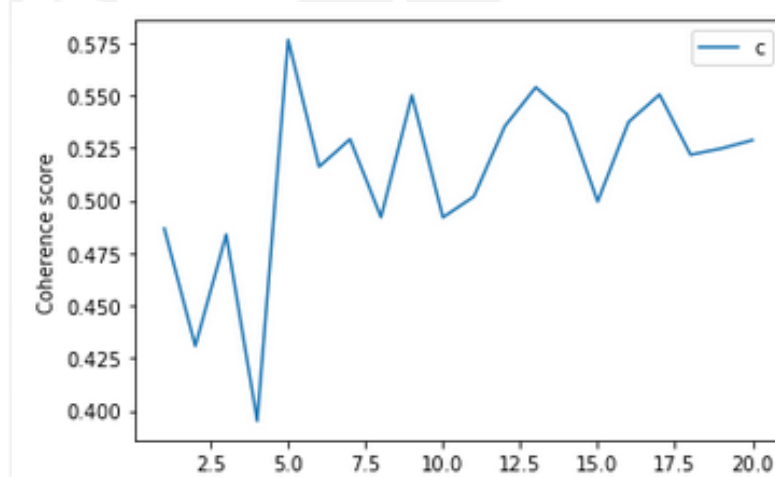
```

Gambar 4.14 Hasil TF-IDF

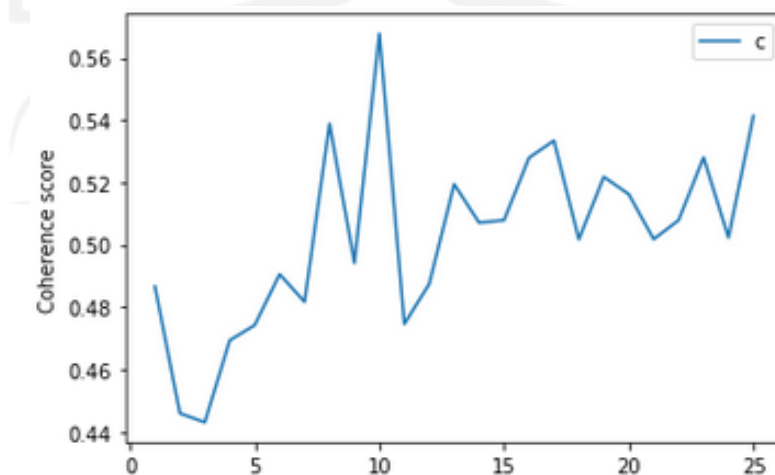
Berdasarkan gambar Gambar 4.14 misalnya (143,1) maka 143 mengartikan kode unik suatu kata misalnya kata “bahaya” dan 1 mempunyai arti kemunculan kata tersebut.

4.4 Nilai koherensi

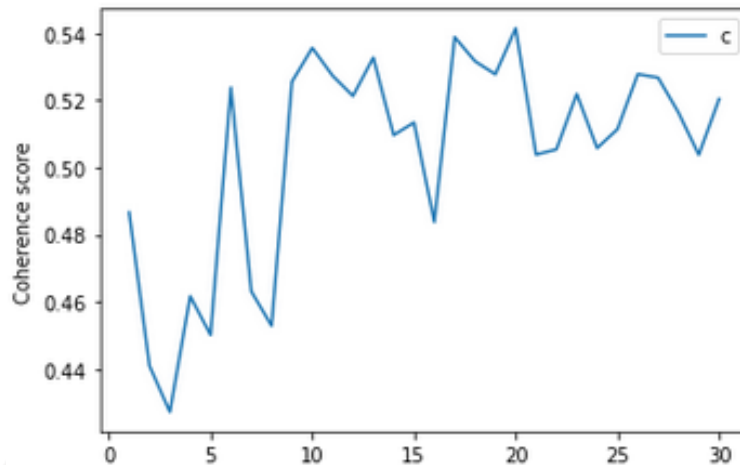
Nilai koherensi dalam pemodelan topik digunakan untuk mengukur seberapa dapat ditafsirkan topik tersebut bagi manusia. Nilai koherensi digunakan untuk membantu didalam menentukan berapa banyak topik yang tepat dari dokumen yang dimiliki. Semakin tinggi nilai koherensi maka semakin bagus hasil interpretasi topik modeling. Pada penelitian ini mencoba mencari nilai terbaik dengan melakukan sebanyak tiga kali percobaan dengan tiap-tiap percobaan mempunyai limit atau batas maksimal jumlah topik yang dilakukan. Hasil percobaan dapat dilihat pada Gambar 4.15 sampai dengan Gambar 4.17. Berdasarkan percobaan yang dilakukan tiap percobaan menghasilkan jumlah topik terbaik yang berbeda-beda.



Gambar 4.15 Jumlah Topik Berdasarkan Nilai Koherensi ke 1



Gambar 4.16 Jumlah Topik Berdasarkan Nilai Koherensi ke 2



Gambar 4.17 Jumlah Topik Berdasarkan Nilai Koherensi ke 3

Berdasarkan pada Tabel 4.2 percobaan ke satu hingga ke tiga mempunyai masing-masing nilai koherensi terbaik dengan membandingkan hasil skor koherensi tiap jumlah topiknya. Pada percobaan ke satu mempunyai jumlah topik terbaik yaitu lima dengan jumlah skor 0.576453. Kemudian percobaan berikutnya jumlah topik terbaik adalah sepuluh dengan skor 0.567734. Percobaan terakhir jumlah topik terbaik adalah dua puluh dengan skor 0.541526. Perbandingan hasil skor percobaan ke satu dengan ke tiga hampir saling mendekati. Pada percobaan ke tiga, klaster yang dihasilkan dalam distribusi kata tiap topik memiliki tingkat kedekatan yang cenderung lebih kecil atau kemiripan dalam distribusi kata tiap topik rendah. Sehingga penelitian ini memutuskan untuk menggunakan hasil dari percobaan pertama dengan mempertimbangkan tingginya skor koherensi dan hal ini mempunyai arti bahwa jumlah topik yang sesuai dalam dokumen yang dimiliki berjumlah lima. Beberapa percobaan dan hasil nilai koherensi pada penelitian ini dapat dilihat pada Tabel 4.2.

Tabel 4.2 Nilai Koherensi

Percobaan ke	1		2		3	
Jumlah Limit	20		25		30	
	Jumlah Topik	Nilai Koherensi	Jumlah Topik	Nilai Koherensi	Jumlah Topik	Nilai Koherensi
	1	0.486593	1	0.486593	1	0.486593
	2	0.430761	2	0.445928	2	0.440847
	3	0.483805	3	0.443049	3	0.427128
	4	0.395044	4	0.469351	4	0.461644

	5	0.576453	5	0.474206	5	0.450067
	6	0.516031	6	0.49053	6	0.523809
	7	0.529143	7	0.481683	7	0.46311
	8	0.491941	8	0.538842	8	0.452798
	9	0.50183	9	0.494138	9	0.52556
	10	0.535317	10	0.567734	10	0.53566
	11	0.50183	11	0.474582	11	0.52735
	12	0.535317	12	0.487454	12	0.521331
	13	0.553851	13	0.519422	13	0.532772
	14	0.541085	14	0.507063	14	0.509668
	15	0.499501	15	0.507892	15	0.513399
	16	0.537353	16	0.527891	16	0.483761
	17	0.550389	17	0.533392	17	0.53888
	18	0.521766	18	0.501728	18	0.531672
	19	0.524718	19	0.521774	19	0.527787
	20	0.528706	20	0.51622	20	0.541526
Jumlah Topik Terbaik	5		10		20	

4.4.1 Pemodelan Topik dengan LDA

Setelah memiliki semua yang diperlukan untuk melatih model LDA seperti *corpus*, *dictionary*, jumlah topik, alpha dan eta. Alpha dan eta merupakan hyperparameters yang akan mempengaruhi *sparsity* topik. Berdasarkan dokumen gensim pemakaian secara *default* keduanya adalah 1.0 dan pada pemodelan LDA ini menggunakan aturan *default* yang ada di gensim. Berikut pada Gambar 4.18 merupakan implementasi kode program LDA.

```
model = LdaModel(corpus=corpus_tfidf, id2word=dictionary, num_topics=5)
for idx, topic in model.print_topics(-1):
    print('Topic: {} Word: {}'.format(idx, topic))
```

Gambar 4.18 LDA

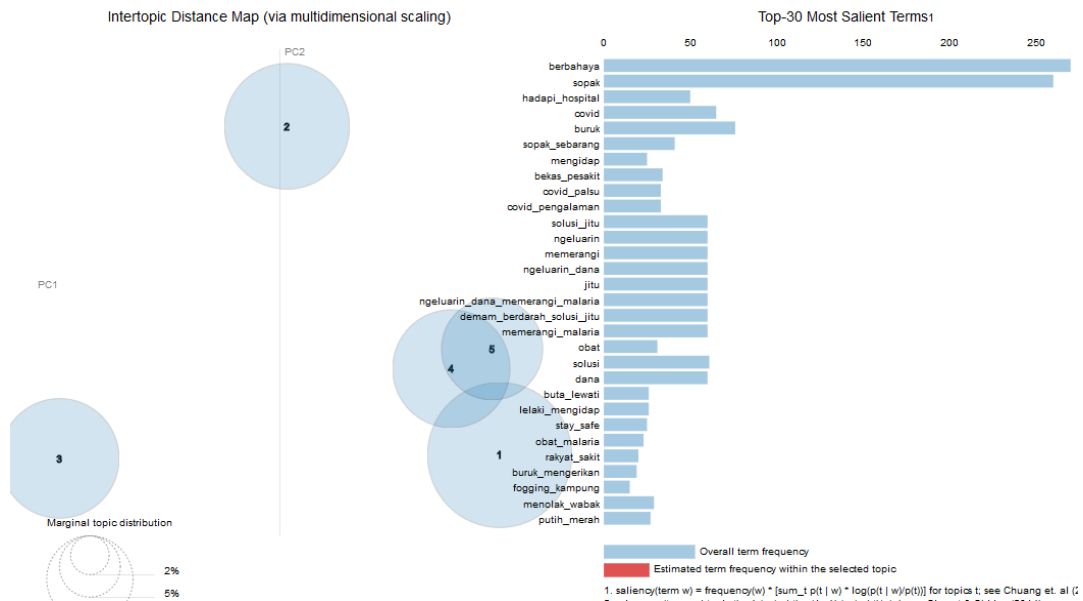
Hasil pemodelan topik menggunakan LDA dapat dilihat pada Tabel 4.3. Kemudian tiap kata dalam hasil pemodelan topik akan mempunyai nilai bobot. Nilai bobot tersebut merupakan suatu nilai peluang kemunculan dari tiap kata di dalam topik.

Tabel 4.3 Hasil Pemodelan Topik

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"
T3	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"
T5	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"

4.5 Visualisasi

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 seperti pada Gambar 4.19.



Gambar 4.19 Visualisasi *Intertopic Distance Map*

Pada Gambar 4.19 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 ke 3 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*.

Visualisasi selanjutnya dengan menggunakan *wordcloud*. Visualisasi *wordcloud* digunakan untuk mempermudah memperoleh wawasan dari topik-topik yang dihasilkan oleh pemodelan (Naury et al., 2021). Visualisasi ini berdasarkan hasil pemodelan LDA pada Gambar 4.19 sehingga terdapat 5 visualisasi *wordcloud* yang menunjukkan hasil pemodelan topik. Distribusi kata tiap topik dimana besar kecilnya kata mempunyai arti terhadap tingkat kemunculan kata. Adapun visualisasi *wordcloud* dapat dilihat pada Gambar 4.20 sampai Gambar 4.24.



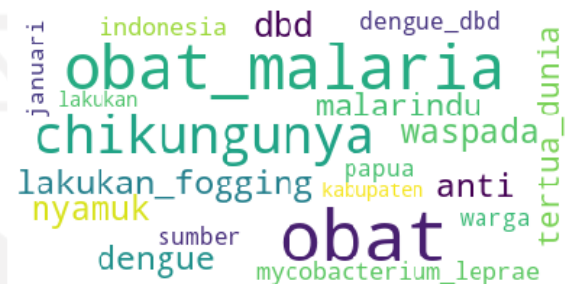
Gambar 4.20 Wordcloud1



Gambar 4.21 Wordcloud2



Gambar 4.22 Wordcloud3



Gambar 4.23 Wordcloud4



Gambar 4.24 Wordcloud5

4.6 Hasil dan Analisa

Berdasarkan hasil pemodelan topik maka penulis akan menganalisis secara kualitatif. Hasil pemodelan topik akan ditafsirkan makna yang terkandung dalam kata-kata yang muncul di tiap topik. Penafsiran makna dari suatu topik diambil berdasarkan 10 kata yang paling dominan seperti yang ditunjukkan pada Tabel 4.4.

Tabel 4.4 Analisa Hasil Pemodelan Topik

Topik	Terms
Topik #1: Dana untuk memerangi malaria dan demam berdarah	Sopak_sebarang, ngeluarin_dana, ngeluarin, solusi_jitu, solusi, memerangi, dana, ngeluarin_dana_memerangi_malaria, demam_berdarah_solusi_jitu, memerangi_malaria
Topik #2: Covid-19	Buruk, covid, sopak, hadapi_hospital, bekas_pesakit, covid_palsu, covid_pengalaman, stay_safe, hospital, berbahaya

Topik	Terms
Topik #3: Buta dan kusta	Berbahaya, sopak, buta_lewati, lelaki_mengidap, rakyat_sakit, sakit, rumah_sakit, penderita, tangan_kakinya_buntung_matanya, belang
Topik #4: Pengobatan dan penanggulangan	Obat, obat_malaria, chikungunya, lakukan_fogging, dbd, dengue, anti, waspada, nyamuk, malarindu
Topik #5: Dana untuk memerangi malaria dan demam berdarah	Mengidap, memerangi_malaria, memerangi, ngeluarin_dana_memerangi_malaria, demam_berdarah_solusi_jitu, ngeluarin, solusi_jitu, jitu, 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. Berdasarkan topik tersebut dapat ditafsirkan bahwa masyarakat Indonesia mengalami kekhawatiran akan penyakit tersebut karena penyakit tersebut tergolong menular dan mematikan. Berdasarkan informasi dari berita yang beredar mengatakan bahwa pada bulan Januari 2021 tingkat kasus demam berdarah meningkat di berbagai daerah khususnya di pulau jawa (Andri, 2021) dan selain itu Indonesia juga tengah menghadapi virus covid-19, Menteri Keuangan Sri Mulyani mengatakan bahwa terdapat alokasi dana untuk penanganan covid-19 sebesar 61,84 triliun (Putri, 2021). Hal tersebut mengartikan bahwa fakta dilapangan, masyarakat Indonesia bukan hanya menghadapi covid-19 saja akan tetapi terdapat penyakit lain yang mematikan yaitu malaria dan demam berdarah. Masyarakat khawatir akan tidak ada bantuan atau pendanaan untuk memerangi penyakit ini karena pemerintah fokus terhadap penanganan covid-19 dan selain itu di tengah wabah banyak masyarakat yang tidak bekerja karena dirumahkan kemudian banyak pekerjaan yang berdampak secara langsung maupun tidak langsung.

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*. 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. Informasi ini dikuatkan dengan adanya berita terkait virus tersebut menurut kementerian kesehatan dalam publikasinya pada bulan januari 2021 terjadi kasus baru

sebanyak 335.116 orang yang dinyatakan sembuh hanya 262.124 orang dan kasus aktif mencapai 65.132 orang (Montesori, 2021).

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. Informasi ini dikuatkan dengan diperingatinya hari kusta sedunia yang jatuh pada tanggal 31 Januari. Hingga saat ini Indonesia masih belum mendapat status eliminasi kusata dan direktur Pencegahan dan Pengendalian Penyakit Menular Langsung (P2PML) kemenkes RI dr. Siti Nadia Tarmizi mengatakan ada 8 provinsi yang belum berhasil mendapatkan status eliminasi kusta (Fundrika & Afrianti, 2021). 8 provinsi tersebut yakni seperti Sumatra Utara, Sulawesi Selatan, Sulawesi Barat, Gorontalo, Maluku, Maluku Utara, Papua dan Papua Barat.

Kemunculan 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 memerlukan obat untuk penyakit malaria dan demam berdarah serta penanggulangan terhadap penyakit tersebut dengan salah satunya melakukan penyemprotan (*fogging*). Kasus demam berdarah pada bulan Januari 2021 meningkat di beberapa wilayah di pulau jawa. Menurut sekretaris dinkes Nia Aminah, Kota Bekasi terdapat 29 kasus (Waluyo, 2021) dan di wilayah Jawa Timur kepala dinkes Herlin Ferliana mengatakan ada 408 kasus (Andri, 2021). Kepala dinkes jatim dan sekretaris dinkes kota bekasi juga menghimbau kepada masyarakat luas untuk mencegah atau menanggulangi kasus penyakit tersebut seperti melakukan gerakan satu rumah satu jumantik dan juga meminta melakukan pemberantasan sarang nyamuk dengan cara 3M yakni menguras, menutup dan mengubur di lingkungan sekitar rumah, kantor, sekolah atau tempat-tempat umum.

Dari hasil pemodelan topik tersebut dapat digunakan oleh pemerintah dalam membantu melihat kondisi atau fakta dilapangan. Hal ini dikuatkan dengan analisis terhadap hasil pemodelan topik pada topik 1 dan topik 5. Dan juga dapat digunakan oleh tenaga kesehatan atau pelayanan kesehatan dalam memberikan promosi ataupun gerakan apa saja yang harus dilakukan oleh masyarakat dalam menanggulangi peningkatan kasus penyakit

dan juga dapat digunakan untuk melihat kebutuhan masyarakat dalam menanggulangi penyakit misalnya penyemprotan *fogging*.

4.7 Pengujian

Tahap pengujian digunakan untuk mengevaluasi kekurangan yang ada dalam penelitian yang sudah dilakukan. Pengujian penelitian ini menggunakan *quisioner* dengan responden peneliti, tenaga kesehatan dan akademis. Dalam pengujian yang diuji berupa pengujian fungsional terhadap hasil pemodelan topik. Responden peneliti dan akademis diisi oleh lingkungan akedemis STMIK Sinar Nusantara dan Universitas Islam Indonesia sedangkan responden kesehatan di isi oleh bidang keperawatan Rumah sakit UNS Surakarta. Jumlah total responden adalah 32 orang. Hasil penilaian pengujian responden dan instrumen yang dibangun dapat dilihat pada Tabel 4.5.

Tabel 4.5 Hasil Pengujian Kuesoner

No.	Pertanyaan	Hasil			
		SB	B	C	K
1	Apakah pemodelan topik pada cuitan tentang penyakit tropis di Indonesia dengan metode <i>Latent Dirichlet Allocation</i> (LDA) mudah dipahami?	11	20	1	0
2	Apakah pemodelan topik pada cuitan tentang penyakit tropis di Indonesia dengan metode <i>Latent Dirichlet Allocation</i> (LDA) memudahkan para pengguna melakukan analisis penelitian topik?	13	17	2	0
3	Apakah pemodelan topik pada cuitan tentang penyakit tropis di Indonesia dengan metode <i>Latent Dirichlet Allocation</i> (LDA) yang dihasilkan sudah sesuai?	12	17	3	0
4	Apakah pemodelan topik pada cuitan tentang penyakit tropis di Indonesia dengan metode <i>Latent Dirichlet Allocation</i> (LDA) bermanfaat bagi peneliti, tenaga kesehatan dan akademis?	14	17	1	0
5	Apakah sistem yang disediakan dapat menunjang untuk mengembangkan penelitian informatika dan kesehatan di Indonesia?	14	18	0	0
	Total Pengujian Responden	64	89	7	0

Keterangan:

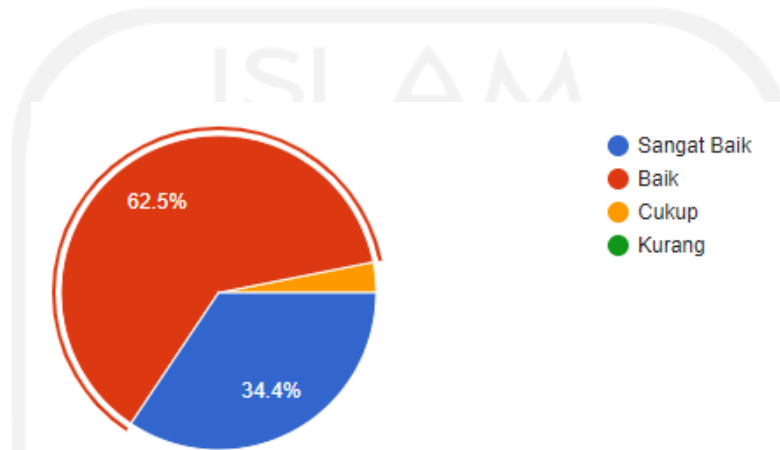
SB : Sangat Baik

B : Baik

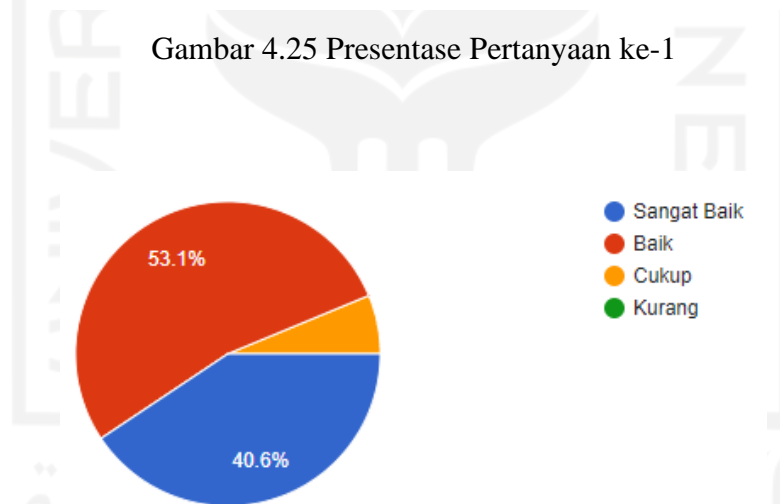
C : Cukup

K : Kurang

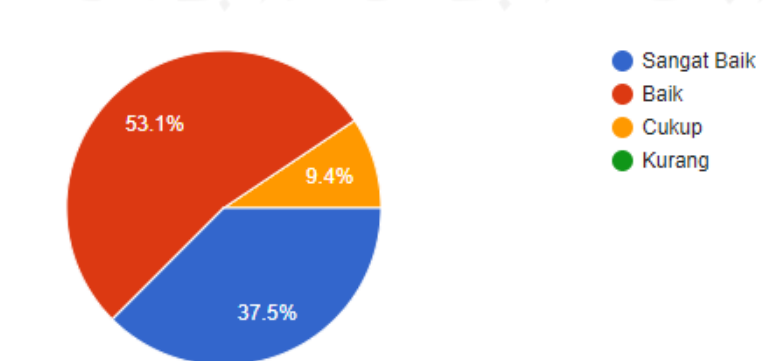
Adapun gambar presentase tiap pertanyaan dapat dilihat pada Gambar 4.25 sampai Gambar 4.29.



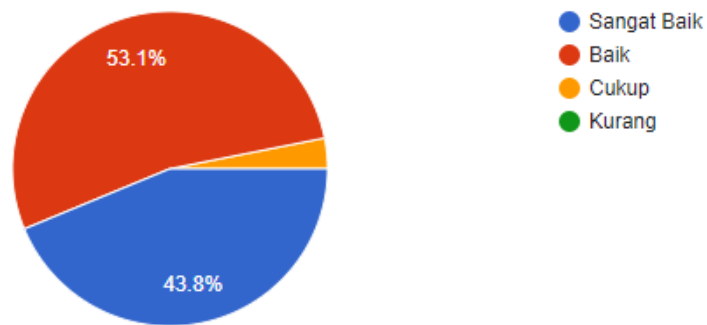
Gambar 4.25 Presentase Pertanyaan ke-1



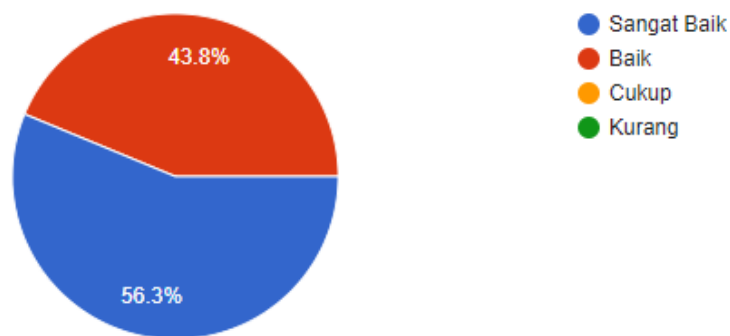
Gambar 4.26 Presentase Pertanyaan ke-2



Gambar 4.27 Presentase Pertanyaan ke-3



Gambar 4.28 Presentase Pertanyaan ke-4



Gambar 4.29 Presentase Pertanyaan ke-5

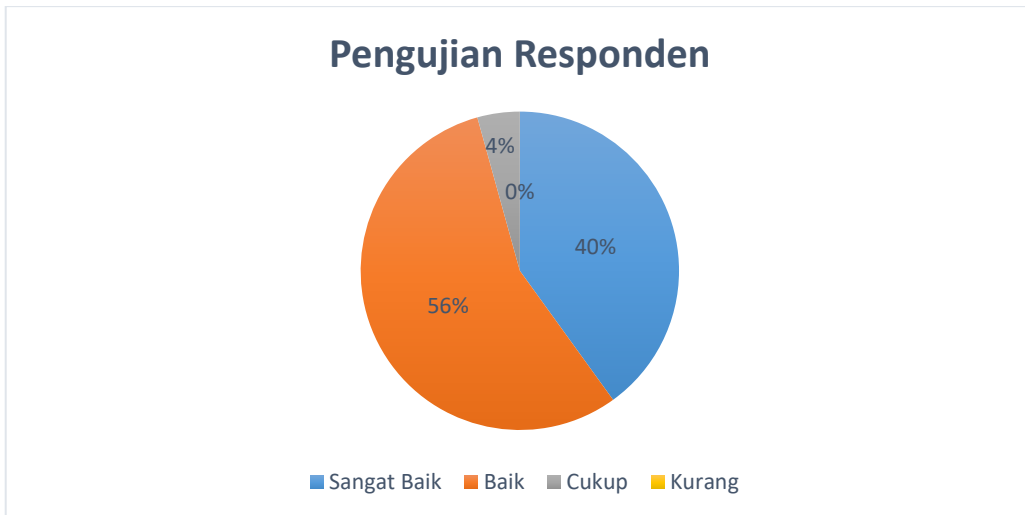
Nilai total keseluruhan pengujian quesoner adalah 160. Responden yang melakukan pengujian memberikan tanggapan sangat baik 40%, baik 55.6% dan cukup 4.3% dengan menggunakan perhitung seperti berikut:

$$\text{Sangat baik} = \frac{64}{160} \times 100\% = 40\%$$

$$\text{Baik} = \frac{89}{160} \times 100\% = 55.6\%$$

$$\text{Cukup} = \frac{7}{160} \times 100\% = 4.3\%$$

Adapun dari hasil perhitungan tersebut disajikan pada persentase Gambar 4.30.



Gambar 4.30 Hasil Pengujian Responden

Berdasarkan hasil pengujian responden yang ditunjukkan pada Gambar 4.30 maka fungsionalitas hasil dari penelitian Pemodelan Topik pada Cuitan tentang Penyakit Tropis di Indonesia dengan Metode *Latent Dirichlet Allocation* dapat dikatakan sudah berjalan dengan baik dengan tanggapan responden sebesar 56% dan 40% menyatakan sangat baik.

BAB 5

Kesimpulan

5.1 Kesimpulan

Penelitian 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 5, berdasarkan hasil nilai koherensi tertinggi yakni 0.576453.

Hasil pemodelan topik dengan LDA dapat ditampilkan dalam bentuk visualisasi *intertopic distance map* dan *wordcloud*. Visualisasi tersebut dapat memberikan gambaran terkait dengan sebaran kata-kata tiap topik yang dihasilkan oleh pemodelan topik dengan LDA. Kedua visualisasi ini dapat dikatakan representatif dalam menampilkan hasil pemodelan topik. Dalam visualisasi *interstopik distance map* hasil pemodelan topik dapat diketahui jarak kedekatan antar topik dan frekuensi kemunculan kata di tiap topiknya sedangkan visualisasi *wordcloud* dapat diketahui dengan frekuensi kemunculan kata dengan digambarkan besar kecilnya huruf.

Berdasarkan penelitian yang telah kami lakukan, hasil pemodelan topik dengan LDA dapat disimpulkan bahwa *trends* 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 55.6% baik, sangat baik 40%.

5.2 Saran

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

Daftar Pustaka

- Alamudi, M. Y. (2018). *Pentingnya Eradikasi Penyakit Tropis Di Indonesia*. (June).
- Alyusi, S. D. (2016). *Media Sosial: Interaksi, Identitas dan Modal Sosial (Pertama)*. Jakarta: Kencana.
- Amelia, R. (2020). *Pendampingan Masyarakat Mantan Penderita Kusta*. Jakarta: Sahabat Alter Indonesia.
- Amiruddin, M. D. (2001). *Penyakit Kusta*. Hasanuddin University Press.
- Amiruddin, M. D. (2012). *Penyakit Kusta Sebuah Pendekatan Klinis* (A. Wijaya, Ed.). Surabaya: Brilian Internasional.
- Andri. (2021). Melawan Covid-19, Mencegah dbd. Retrieved January 10, 2022, from halojatim website: <https://halojatim.com/read/melawan-covid-19-mencegah-dbd>
- 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.
- Duarsa, A. B. S. (2020). *Epidemiologi Penyakit Menular: Studi Ekologi Malaria*. Malang: Media Nusa Creative.
- Fundrika, B. A., & Afrianti, D. (2021). Hari Kusta Sedunia 2021: Ini 8 Provinsi yang Belum Berhasil Eliminasi. Retrieved January 10, 2022, from suara.com website: <https://www.suara.com/health/2021/01/29/182500/hari-kusta-sedunia-2021-ini-8-provinsi-yang-belum-berhasil-eliminasi?page=all>
- Hidayatullah, A. F., Aditya, S. K., Karimah, & Gardini, S. T. (2019). Topic modeling of weather and climate condition on twitter using latent dirichlet allocation (LDA). *The International Conference on Information Technology and Digital Applications*. <https://doi.org/10.1088/1757-899X/482/1/012033>
- Ipmawati, J., Kusriani, & Luthfi, E. Ta. (2017). Komparasi Teknik Klasifikasi Teks Mining Pada Analisis Sentimen. *Ijns*, 6(1), 28–36.
- Irwan. (2017). *Epidemiologi Penyakit Menular*. Yogyakarta: CV. Absolute Media.
- Jelodar, H., & Wang, Y. (2018). *Latent Dirichlet Allocation (LDA) and Topic modeling: models, applications, a survey*.
- Kang, H. J., Kim, C., & Kang, K. (2019). *Analysis of the Trends in Biochemical Research Using Latent Dirichlet Allocation (LDA)*. 1–14.
- Kashyap, R., & Nahapetian, A. (2014). *Tweet Analysis for User Health Monitoring*. 348–351. <https://doi.org/10.4108/icst.mobihealth.2014.257537>

- McClellan, C., Ali, M. M., Mutter, R., Kroutil, L., & Landwehr, J. (2017). *Using social media to monitor mental health discussions 2 evidence from Twitter*. 24(October 2016), 496–502. <https://doi.org/10.1093/jamia/ocw133>
- Montesori, J. (2021). Kasus Covid-19 di Januari 2021 Terburuk Sejak Pandemi Landa Indonesia. Retrieved January 10, 2022, from beritasatu website: <https://www.beritasatu.com/kesehatan/726469/kasus-covid19-di-januari-2021-terburuk-sejak-pandemi-landa-indonesia>
- 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>
- Pratomo, Y. (2019). Untuk Pertama Kali, Twitter Ungkap Jumlah Pengguna Harian. Retrieved December 29, 2021, from kompas.com website: <https://tekno.kompas.com/read/2019/02/09/11340027/untuk-pertama-kali-twitter-ungkap-jumlah-pengguna-harian>
- Putra, I. M. K. B. (2017). *Analisis Topik Informasi Publik Media Sosial Di Surabaya Menggunakan Pemodelan Latent Dirichlet Allocation (LDA)*.
- Putri, C. A. (2021). Anggaran Covid-19 2021 Rp 61,8 T, Sri Mulyani: Bisa Nambah! Retrieved January 10, 2022, from CNBC Indonesia website: <https://www.cnbcindonesia.com/news/20210121141648-4-217781/anggaran-covid-19-2021-rp-618-t-sri-mulyani-bisa-nambah>
- Ratnadewi, Christine, M., Manurung, R. T., Wargasetia, T. L., & Dani. (2019). Pelatihan Pencegahan dan Penanggulangan Demam Berdarah Dengue. *Celebes Abdimas: Jurnal Pengabdian Kepada Masyarakat*, 1(April), 36–44.
- Ruminem, Tandirogang, N., Bakhtiar, R., Rahayu, A. P., & Kadir, A. (2020). *Modul Penyakit tropis* (Sholichin, Ed.). Samarinda: Gunawana Lestari.
- Sidana, S., Amer-yahia, S., Clausel, M., Rebai, M., Mai, S. T., & Amini, M. (2018). *Health Monitoring on Social Media over Time*. 14(8), 1–14.
- Siregar, A. M., & Puspabhuana, A. (2017). *Data Mining: Pengolahan Data Menjadi Informasi dengan RapidMiner*. Surakarta: CV Kekata Group.
- 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=PP1&dq=Penyakit+Tropis+Di+Indonesia&ots=dqRj2tLHBp&sig=QUONUNevl6s555S>

- MPrax2y1x-kw&redir_esc=y#v=onepage&q=Penyakit Tropis Di Indonesia&f=false
- Supranelfy, Y., & Oktarina, R. (2021). Gambaran Perilaku Pencegahan Penyakit Malaria di Sumatera Selatan (Analisis Lanjut Riskesdas 2018). *BALABA*, 17, 19–28.
- Waluyo, D. (2021). 1.665 Warga Bekasi Terserang DBD Sejak Januari hingga Oktober 2021. Retrieved January 10, 2022, from kompas.com website: <https://megapolitan.kompas.com/read/2021/11/12/16441391/1665-warga-bekasi-terserang-dbd-sejak-januari-hingga-oktober-2021>
- 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.
- Yulidar, & Dinata, A. (2016). *Rahasia Daya Tahan Hidup Namuk Demam Berdarah: Cara Cerdas Mengenal Aedes Aegypti dan Kiat Sukses Pengendalian Vektor DBD*. Yogyakarta: Deepublish.

