



Analisis Jejaring Penelitian Kesehatan Indonesia Menggunakan *Text Mining*

Yoga Sahria

17917225

Tesis diajukan sebagai syarat untuk meraih gelar Magister Komputer

Konsentrasi Informatika Medis

Program Studi Informatika Program Magister

Fakultas Teknologi Industri

Universitas Islam Indonesia

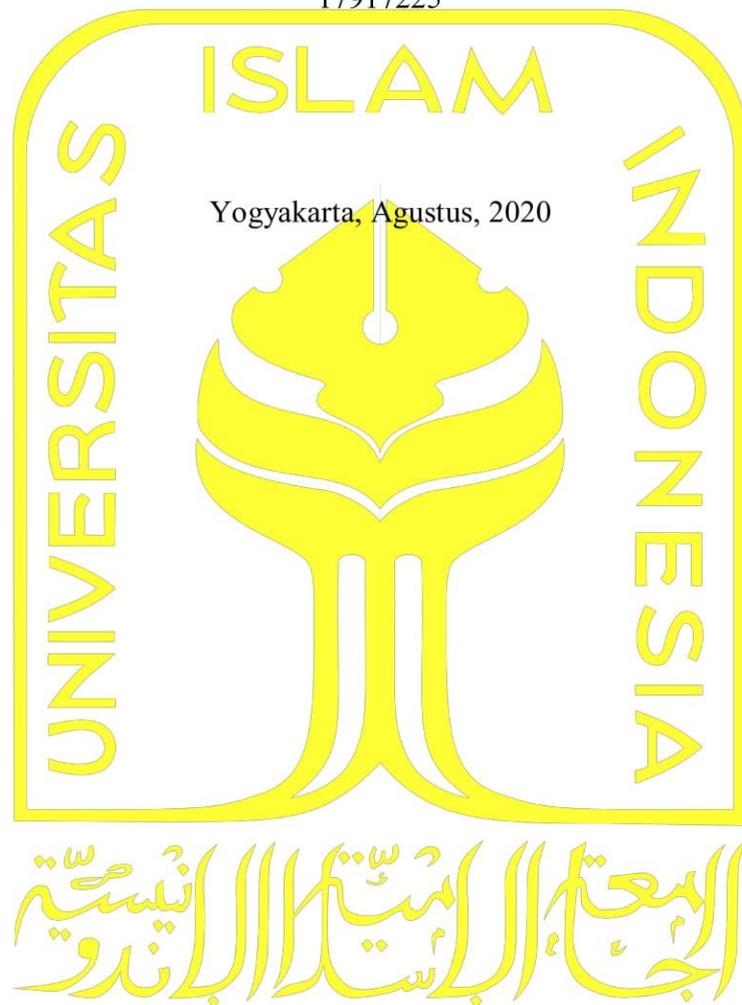
2020

Lembar Pengesahan Pembimbing

Analisis Jejaring Penelitian Kesehatan Indonesia Menggunakan *Text Mining*

Yoga Sahria

17917225



Pembimbing

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

Lembar Pengesahan Penguji

Analisis Jejaring Penelitian Kesehatan Indonesia Menggunakan *Text Mining*

Yoga Sahria

17917225

ISLAM

Yogyakarta, Agustus, 2020

Tim Penguji,

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

Ketua



Ing.Ridho Rahmadi, S.Kom., M.Sc., Ph.D.

Anggota I



Dr. Raden Teduh Dirgahayu, S.T., M.Sc.

Anggota II



Mengetahui,

Ketua Program Studi Informatika Program Magister

Universitas Islam Indonesia



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

Abstrak

Analisis Jejaring Penelitian Kesehatan Indonesia Menggunakan Text Mining

Penyimpanan repositori jurnal penelitian kesehatan di Indonesia semakin banyak, akan tetapi minim riset untuk dianalisis dan dimodelkan. Penelitian ini mengusulkan sebuah basis pemodelan pengetahuan dengan *graph*. Peneliti menggunakan data yang diambil dari website SINTA khusus di bidang kesehatan dengan *keyword* kesehatan dan *health*. Data yang diperoleh dengan menggunakan teknik web *scraping*. Tujuan dari penelitian ini yaitu membangun pengetahuan jejaring penelitian kesehatan Indonesia berdasarkan *author*, judul, topik, afiliasi, dan tahun dengan metode *text mining*, memvisualisasikan penelitian kesehatan Indonesia berdasarkan *author*, judul, topik, afiliasi, tahun dengan *graph database* dan menganalisis tren penelitian kesehatan Indonesia berdasarkan judul penelitian. Basis pemodelan pengetahuannya menggunakan *graph* yang kemudian menerapkan perintah *query* bertujuan untuk melihat jejaring penelitian di Indonesia berdasarkan *author*, judul, topik, afiliasi, dan tahun. Adapun untuk analisis topik menggunakan metode *topic modelling* yang digunakan untuk melihat tren suatu topik judul penelitian pada tahun tertentu. Metode *topic modeling* yang digunakan adalah LDA (*Latent Dirichlet Allocation*). Pemodelan *topic modelling* yang dilakukan tujuannya untuk mengidentifikasi tren topik-topik yang sering muncul terhadap judul-judul penelitian di bidang penelitian kesehatan di Indonesia secara keseluruhan dan pada tahun-tahun tertentu. Hasil penelitian diharapkan mampu memberikan solusi untuk membantu peneliti untuk mempermudah peneliti dalam mempersiapkan penelitiannya khususnya bidang kesehatan di Indonesia.

Kata kunci

Penelitian kesehatan, Jejaring, *text mining*, *graph*

Abstract

Indonesian Medical Research Network Analysis Through Text Mining

There are more and more repositories of medical research journals in Indonesia, but there is minimal research to analyze and model. This study proposes the basis for modeling knowledge with graphs. Researchers used data taken from the SINTA website. Data obtained using web erosion techniques. The purpose of this research is to build knowledge of the Indonesian health research network based on the author, title, topic, relationship, and year using the text mining method, visualizing Indonesian health research based on the author, title, topic, relationship, year with database charts and tracking health research trends Indonesia based on the research title. The development of the knowledge base model uses graphics which then apply query queries that are used to view research networks in Indonesia based on the author, title, topic, affiliation, and year. As for the topic analysis, the topic modeling method is used to see the research topic trends in a particular year. The topic modeling method used is LDA (Latent Dirichlet Allocation). Topic modeling is done for modeling topics on research titles in the field of overall health research and certain years. The results of the study are expected to provide a solution for researchers to facilitate researchers in preparing their research specifically in the field of health in Indonesia.

Keywords

Health research, networking, text mining, *graphs*

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.

Yogyakarta, 22 Agustus, 2020



Yoga Sahria, S.Kom

Daftar Publikasi

Publikasi Jurnal

1. Pemodelan Pengetahuan *Graph* Database Untuk Jejaring Penelitian Kesehatan di Indonesia (Format Referensi IEEE)
(Jurnal Media Informatika Budidarma SINTA 3)
DOI: <http://dx.doi.org/10.30865/mib.v4i3.2183>
2. *Analysis of Health Research Topics in Indonesia Using the LDA (Latent Dirichlet Allocation) Topic Modeling Method* (Format Referensi IEEE)
(Jurnal RESTI SINTA 2) DOI: <https://doi.org/10.29207/resti.v4i2.1821>

Publikasi *Proceeding*

1. Implementasi Teknik *Web Scraping* pada Jurnal SINTA Untuk Analisis Topik Penelitian Kesehatan Indonesia (Format Referensi Vancouver)
(*Proceeding of The 11th University Research Colloquium 2020: Bidang Sains dan Teknologi*)

Publikasi yang menjadi bagian dari tesis

Publikasi berikut menjadi bagian dari Bab 3

Sitasi publikasi 1

Kontributor	Jenis Kontribusi
Yoga Sahria	Mendesain eksperimen (60%) Menulis <i>paper</i> (100%)
Dhomas Hatta Dudholi, Ph.D	Konsep eksperimen (40%) Review <i>paper</i> (50%)

Sitasi publikasi 2

Kontributor	Jenis Kontribusi
Yoga Sahria	Mendesain eksperimen (80%) Menulis dan mengedit <i>paper</i> (80%)
Dhomas Hatta Fudholi, Ph.D	Mendesain eksperimen (20%) Review <i>paper</i> (20%)

Sitasi Publikasi 3

Kontributor	Jenis Kontribusi
Yoga Sahria	Mendesain eksperimen (100%) Menulis <i>paper</i> (100%)

Halaman Kontribusi

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

1. Bapak DThomas Hatta Fudholi, S.T., M.Eng., Ph.D. yang berkontribusi sebagai pembimbing tesis secara penuh.
2. Bapak Ahmad Munasir Raf'ie Pratama, S.T, M.I.T., Ph.D sebagai penguji proposal tesis yang memberikan saran serta pendapat sehingga proses penulisan tesis ini menjadi lebih baik dengan masukan-masukan yang telah diberikan.
3. Bapak Dr. Syarif Hidayat S.Kom., M.I.T. sebagai penguji proposal tesis yang memberikan saran serta pendapat sehingga proses penulisan tesis ini menjadi lebih baik dengan masukan-masukan yang telah diberikan.
4. Bapak Ahmad Fathan Hidayatullah S.T., M.Cs. sebagai penguji progress yang memberikan masukan sehingga kekurangan yang terdapat pada penulisan tesis ini dapat diperbaiki sesuai arahan yang telah diberikan.
5. Bapak Taufiq Hidayat S.T., M.Cs. sebagai penguji progress yang memberikan masukan sehingga kekurangan yang terdapat pada penulisan tesis ini dapat diperbaiki sesuai arahan yang telah diberikan.
6. Ing,Ridho Rahmadi, S.Kom., M.Sc., Ph.D sebagai penguji pendadaran yang memberikan masukan sehingga lebih melengkapi kekurangan yang terdapat pada penulisan tesis ini.
7. Dr. RadenTeduh Dirgayahu sebagai penguji pendadaran yang memberikan masukan sehingga lebih melengkapi kekurangan yang terdapat pada penulisan tesis ini.
8. Bu Izzati Muhimmah, S.T., M.Sc., Ph.D. sebagai dosen dan Kaprodi Magister Informatika sekaligus dosen pendamping yang selalu menyemangati.
9. Dosen, Mahasiswa dan Akademik STIKES Al Islam Yogyakarta yang telah memberikan arahan tesis sebagai pakar kesehatan dan pengisis kuisoner.
10. Bapak dr. Wahyudi Santoso sebagai dokter umum yang telah mengisi kuisoner penelitian.
11. Kontribusi Pihak Jurusan Informatika Universitas Islam Indonesia yang memberikan insentif publikasi Ilmiah Terakreditasi SINTA 2.

Halaman Persembahan

Bismillahirrohmanirrohim dengan izin, ridho, kemurahan dan kuasa Allah SWT, Tesis penelitian ini saya persembahkan kepada kedua orang tua ibu Tri Darwati dan bapak Yatmin, adikku tersayang Zulvia Talita Bunda dan keluarga saya, Semoga Allah membalas semua kebaikan yang engkau berikan.

Kata Pengantar

Assalaamu'alaikum warahmatullaahi wabarakatuh.

Alhamdulillahirabbilalamiin. Segala puji kehadiran Allah SAW atas segala nikmat, karunia, hidayah, dan pertolongan-Nya, sehingga TESIS yang berjudul “**Analisis Jejaring Penelitian Kesehatan Indonesia Menggunakan *Text Mining***” ini dapat terselesaikan dengan baik dan sebagaimana mestinya.

Tesis ini dibuat sebagai syarat yang harus direncanakan, ditempuh dan dipenuhi untuk menyelesaikan pendidikan Magister dari Program Studi Informatika program Magister Kosentrasi Informatika Medis, Fakultas Teknologi Industri, Universitas Islam Indonesia.

Dalam proses penyelesaian Tesis ini, Banyak rintangan dan hambatan yang harus dilalui oleh penulis. Akan tetapi, berbagai rintangan dan hambatan tersebut akhirnya dapat dilewati dan diselesaikan karena dukungan dari berbagai pihak. Maka atas hal tersebut, penulis ingin mengucapkan terima kasih kepada:

1. Kedua orang tuaku tercinta, Tri Darwati dan Yatmin, yang senantiasa memberikan semangat, doa dan dukungan yang sangat berarti untuk anakmu ini.
2. Adikku tersayang, Zulvia Talita Bunda.
3. Semua Keluarga Wasio Atmodipuro.
4. Bapak Profesor Fathul Wahid, S.T., M.Sc., Ph.D., selaku Rektor Universitas Islam Indonesia.
5. Bapak Profesor Dr. Ir. Hari Purnomo, M.T., selaku Dekan Fakultas Teknologi Industri, Universitas Islam Indonesia.
6. Bapak Hendrik, S.T., M.Eng., selaku Ketua Jurusan Teknik Informatika, Universitas Islam Indonesia.
7. Ibu Izzati Muhimmah, S.T., M.Sc., Ph.D., selaku Ketua Program Studi Magister Informatika, Universitas Islam Indonesia.
8. Bapak Dhomas Hatta Fudholi, S.T., M.Eng., Ph.D., selaku dosen pembimbing yang telah berkenan membimbing dan memberikan saran serta masukan untuk penulis.
9. Bapak Ahmad Fathan Hidayatullah, S.T., M.Cs., selaku dosen yang telah berkenan memberikan ilmu nya.
10. Semua dosen magister Informatika, khususnya dosen kosentrasi megister Informatika Medis Universitas Islam Indonesia

11. Semua dosen STIKES Al Islam Yogyakarta, selaku review tesis yang saya teliti
12. Semua teman-teman Pondok Pesantren Al Munawir Komplek L Krapyak Yogyakarta
13. Sahabat-sahabatku dan teman-teman semua yang tidak bisa disebutkan satu persatu untu menyelesaikan Tesis ini.
14. Pihak-pihak lain yang telah membantu kelancaran proses Tesis ini yang tidak dapat disebutkan satu persatu.

Dengan selesainya penulisan laporan penelitian tesis ini, besar harapan penulis agar karya ini dapat memberikan manfaat dan dapat dikembangkan di masa yang akan datang.

Yogyakarta, 01 Agustus 2020

(Yoga Sahria)

Daftar Isi

1.1	Pendahuluan	1
1.2	Latar Belakang.....	2
1.3	Permasalahan.....	4
1.4	Rumusan Masalah.....	4
1.5	Batasan Masalah	4
1.6	Tujuan Penelitian.....	5
1.7	Manfaat Penelitian	5
1.8	Metodologi Penelitian	5
1.9	Sistematika Penulisan	7
2.1	Penelitian Sebelumnya.....	8
2.2	Text Mining.....	11
2.3	Web Scraping	13
2.4	Jurnal SINTA.....	15
2.5	Penelitian Kesehatan.....	15
2.6	Metode Social Network	18
2.7	Teori <i>Graph</i>	19
2.8	Pemodelan Topik.....	24
3.1	Studi Literatur.....	29
3.2	Pengumpulan Data.....	30
3.3	Ekstrasi Informasi.....	30
3.4	Pemodelan	31
3.5	Analisis dan Evaluasi	32
4.1	Pengumpulan Data.....	33
4.2	Hasil Pemodelan Jejaring	34
4.3	Hasil Analisis Pemodelan Pengetahuan	36
4.4	Implementasi Pemodelan Topik	42

4.4.1	<i>Preprosesing</i>	43
4.4.2	Identifikasi Frasa	44
4.4.3	Perhitungan Bobot Setiap Kata	45
4.4.4	Pemodelan Topik	46
4.4.5	Visualisasi Pemodelan Topik	48
4.4.6	Pemodelan Topik Pada Tahun-Tahun Tertentu.....	51
4.5	Analisis dan Evaluasi	53
5.1	Kesimpulan.....	59
5.2	Saran	60

Daftar Tabel

Tabel 2.1 Penelitian Terdahulu	8
Tabel 2.2 <i>Graph Database</i> VS RDBMS	10
Tabel 2.3 Sintak pada CQL	12
Tabel 4.1 Cuplikan Hasil Target Nama Jurnal Penelitian Kesehatan di Indonesia.....	33
Tabel 4.2 Contoh Hasil <i>Scraping</i>	34
Tabel 4.3 Hasil Pemodelan Topik Secara Keseluruhan.....	51
Tabel 4.4 Topik-Topik Pada Tahun 2017.....	51
Tabel 4.5 Topik-Topik Pada Tahun 2018.....	52
Tabel 4.6 Topik-Topik Pada Tahun 2019	52
Tabel 4.7 Hasil Pengujian Quisoner	55

Daftar Gambar

Gambar 2.1. Proses <i>Text Mining</i>	1
Gambar 2.2 Langkah <i>Web Scraping</i>	2
Gambar 2.3 Proses <i>Scraping</i>	13
Gambar 2.4 Proses Penelitian	15
Gambar 2.5. <i>Undirected Cyclic Graph</i>	17
Gambar 2.6 Graph tak berarah berlabel	17
Gambar 2.7 Struktur <i>Database Neo4j</i>	20
Gambar 2.8 <i>Node & value</i>	24
Gambar 2.9 Hubungan antar <i>node</i> satu arah & lebih	25
Gambar 2.10 Konsep Pemodelan Topik	26
Gambar 2.11 Proses LDA	27
Gambar 2.12 Cara kerja <i>Latent Direct Allocation</i> LDA	27
Gambar 2.13 Representasi Model LDA	28
Gambar 3.1 Metode Penelitian	28
Gambar 3.2 Langkah Ekstrasi Informasi	30
Gambar 3.3 <i>Networking</i> Analisis Publikasi	31
Gambar 4.1 Membuat <i>Node</i>	35
Gambar 4.2 <i>Query</i> antar <i>Node</i>	35
Gambar 4.3 Hasil Pemodelan Jejaring Penelitian Kesehatan di Indonesia.....	35
Gambar 4.4 Perbesar Hasil Pemodelan.....	36
Gambar 4.5 Topik – Topik Berdasarkan Tahun.....	37
Gambar 4.6 Topik-Topik Penelitian Kesehatan Pada Tahun 2018.....	38
Gambar 4.7 Topik-Topik Penelitian Kesehatan Pada Tahun 2019.....	38
Gambar 4.8 Topik-Topik Penelitian Kesehatan Pada Tahun 2017.....	39
Gambar 4.9 Topik Riset Malaria dan Afiliasinya.....	40
Gambar 4.10 Penelitian Topik Malaria dan Author.....	41
Gambar 4.11 Judul Penelitian yang berkaitan dengan Malaria	42
Gambar 4.12 Cuplikan kode tahap menghapus data.....	43
Gambar 4.13 Cuplikan kode tahap menghapus angka.....	43
Gambar 4.14 Cuplikan kode <i>Case Folding</i>	44
Gambar 4.15 Cuplikan kode menghapus kalimat kurang dari 3 kata.....	44
Gambar 4.16 Cuplikan kode <i>stopwords</i>	44

Gambar 4.17 Tahap Identifikasi Frasa.....	45
Gambar 4.18 Kode model bigram dan trigram.....	45
Gambar 4.19 Kode program untuk membuat <i>dictionary</i>	45
Gambar 4.20 Cuplikan Kode Program <i>bag of words</i>	46
Gambar 4.21 Kode program untuk menghitung nilai TF-IDF.....	46
Gambar 4.22 Cuplikan Kode Pemodelan Topik LDA.....	47
Gambar 4.23 Nilai Koherensi.....	47
Gambar 4.24 Nilai Topik Koherensi.....	48
Gambar 4.25 Cuplikan kode menyimpan model dalam bentuk pyLDAvis.....	48
Gambar 4.26 Hasil Visualisasi pyLDAvis.....	49
Gambar 4.27 Cuplikan kode program untuk membuat dan menyimpan word cloud.....	50
Gambar 4.28 Visualiasi word cloud.....	50
Gambar 4.29 Peta Persebaran Penyakit Malaria Indonesia.....	54
Gambar 4.30 Prosentase Pertanyaan kuisioner nomor 1.....	56
Gambar 4.31 Prosentase Pertanyaan kuisioner nomor 2.....	56
Gambar 4.32 Prosentase Pertanyaan kuisioner nomor 3.....	56
Gambar 4.33 Prosentase Pertanyaan kuisioner nomor 4.....	56
Gambar 4.34 Prosentase Pertanyaan kuisioner nomor 5.....	57
Gambar 4.35 Hasil Pengujian Responden.....	58

Glosarium

CQL	- Chyper <i>Query</i> Language
SINTA	- Science and Technology Index
LDA	- Latent Dirichlet Allocation
API	- Aplication Programming Interface
RDBMS	- Relational Database Management System
SNA	- Social Network Analysis
ADT	- Abstract data type

BAB 1

Pendahuluan

1.1 Pendahuluan

Penelitian kesehatan adalah suatu langkah metode atau karya ilmiah yang berorientasikan atau memfokuskan pada kegiatannya pada masalah-masalah yang timbul di bidang kesehatan (Ministry of Health Republic of Indonesia 2016). Kesehatan terdiri dari dua sub bidang pokok, yakni yang pertama kesehatan individu berorientasikan klinis dan pengobatan. Penelitian kesehatan pada sub bidang kedua yaitu berorientasi pada kelompok atau masyarakat, yang bersifat pencegahan. Selanjutnya sub bidang kesehatan terdiri dari berbagai disiplin ilmu, seperti informatika medis, kedokteran, keperawatan, epidemiologi, pendidikan kesehatan, kesehatan lingkungan, manajemen pelayanan kesehatan, gizi dan lain-lain. Pada sub-sub bidang tersebut saling berkaitan dan dapat mempengaruhi kondisi kesehatan masyarakat pada umumnya.

Semakin pesat dan perkembangan teknologi penelitian kesehatan di Indonesia banyak dilakukan oleh peneliti untuk melakukan penelitian tema di bidang kesehatan. Semakin banyaknya penelitian media penyimpanan digital telah mendorong terjadinya banyak jumlah dokumen elektronik yang tersimpan dalam *repository* jurnal atau publikasi ilmiah (Barat et al., n.d. 2018). Berbagai karya ilmiah bersumber dari sivitas akademika seperti artikel, tesis, disertasi, skripsi, laporan penelitian, laporan kerja praktek dan lain sebagainya telah tersedia dalam versi digital. Namun, pada umumnya fenomena hal ini tidak disertai dengan pertumbuhan jumlah analisis informasi atau pengetahuan yang dapat disarikan dari dokumen karya ilmiah tersebut.

Pada umumnya publikasi yang disampaikan dalam jurnal tersebut terdiri atas beberapa topik seperti jurnal politik, olahraga, ekonomi, kesehatan, dan lain-lain. Sehingga sangatlah sulit untuk segera dapat mengetahui topik penelitian populer pada tahun tertentu ataupun kecenderungan minat penelitian dalam bidang kesehatan tertentu misalnya. Dalam penelitian ini akan difokuskan pada penelitian kesehatan di Indonesia untuk memetakan jurnal atau publikasi tersebut dapat terpetakan berdasarkan Judul, afiliasi, *author*, topik dan tahun sehingga membantu para pengambil kebijakan peneliti ataupun akademis untuk dapat dengan cepat menganalisis tren topik penelitian kesehatan tertentu yang akan diteliti. Adapun beberapa kemanfaatan strategis yang didapat melalui analisis tentang tren topik

adalah sebagai petunjuk untuk menyusun rencana penelitian kesehatan yang sedang tren, dan sebagai bahan evaluasi untuk menentukan strategi pengembangan penelitian dibidang kesehatan yang berguna yang dapat membantu masyarakat indonesia, Sebagai bahan kajian untuk pengembangan penelitian kesehatan di Indonesia.

Metode *text mining* adalah sebuah metode yang mengembangkan dari metode *data mining* kemudian dapat diterapkan untuk mengatasi masalah-masalah yang sudah diuraikan tersebut (Jo 2018). Algoritma-algoritma dalam *text mining* dibuat untuk dapat mengenali data yang sifatnya terstruktur memetakan jurnal atau publikasi ilmiah tersebut berdasarkan afiliasi dan topik yang terdapat di website jurnal yang berkaitan dengan penelitian kesehatan di Indonesia yang di *scraping* dari dokumen-dokumen yang terdapat di jurnal/website publikasi ilmiah. Sehingga sangatlah sulit untuk segera dapat mengetahui topik penelitian populer pada tahun tertentu.

1.2 Latar Belakang

Dengan perkembangan zaman yang begitu pesat disertai dengan teknologi yang semakin maju semakin banyak peneliti untuk riset yang berfokus dalam bidang kesehatan. Penelitian kesehatan adalah langkah metode ilmiah yang memfokuskan atau berorientasikan pada masalah-masalah yang timbul di dalam bidang kesehatan. Penelitian kesehatan dalam rangka dalam mengatasi memecahkan sebuah permasalahan di bidang kesehatan dengan berbagai pengaruh. Penelitian kesehatan memiliki dua sasaran yaitu memecahkan masalah kesehatan individu yang mengalami kendala kesehatan dan berorientasi pada kelompok atau masyarakat agar dapat mempertahankan kesehatan agar tetap sehat. Penelitian kesehatan merupakan hal yang penting untuk mendukung pengembangan ilmu pengetahuan dibidang kesehatan (Rahmi Surayya 2018). Penelitian kesehatan tersebut kemudian salah satunya disimpan dalam sebuah OJS (*Open Jurnal System*) yang dapat diakses oleh semua kalangan (Sistem, Sahria, and Fudholi 2021). Hal ini menyebabkan banyak penelitian kesehatan di Indonesia banyak tersimpan dalam *repository* Jurnal.

Penelitian terkait dengan menggunakan *graph database* banyak dilakukan oleh peneliti diberbagai domain. Penelitian yang dilakukan (Wirawan and Riyanto 2017) menggunakan *graph database* pada domain transportasi tujuannya untuk memodelkan rute transportasi dan mengetahui perpindahan koridor bus rapid transit. Penelitian selanjutnya (Lu, Hong, and Shi 2017) dengan menerapkan *graph database* pada domain film pemodelan pengetahuan yang dikaji yaitu untuk menganalisis data film. Hasil dari penelitian tersebut menganalisis pengetahuan antara sutradara, aktor pada data film.

Penelitian yang dilakukan oleh (Zhao, Hong, and Shi 2019) mampu menerapkan *graph database Neo4j* dalam domain kesehatan khusus penyakit. Penelitian ini mampu memodelkan penyakit dari berbagai gejala pasien seperti radang, usus buntu, sakit perut, demam, pernapasan. Hasil dari penelitiannya yaitu dapat menghubungkan antara penyakit dan penyakit lainnya dari sebab gejala sehingga dapat ditampilkan dengan jelas untuk membantu orang lain dalam menilai penyakitnya, memberikan pengobatan yang sesuai dengan penyakit yang diderita dengan menggunakan *query*. Penelitian (Constantinov et al. 2018) juga meneliti terkait tentang pemodelan pengetahuan untuk analisis di media sosial. Penelitian tersebut memodelkan data dengan visualisasi untuk menganalisis jejaring media sosial, pemodelan data dengan menggunakan *graph database* dapat mengetahui dengan mudah dalam kasus sektor industri, dan informasi geografis dalam data di media sosial. Penelitian (Bajaj et al. 2018) mampu memodelkan pengetahuan dalam bidang makanan. Hasil dari penelitian tersebut dapat menyajikan atau merekomendasikan resep makanan dengan mengimplementasikan *graph database Neo4j* berdasarkan ketersediaan bahan pengguna.

Dari penelitian-penelitian yang sudah diuraikan peneliti mengusulkan untuk memodelkan pengetahuan jejaring penelitian kesehatan di Indonesia. Penelitian ini berbeda dengan penelitian sebelumnya yaitu berfokus pada domain *Health Research* di Indonesia. Data yang dihasilkan yaitu dari SINTA dengan *keyword* kesehatan dan *health*. *Graph database* dapat memberikan solusi alternatif untuk menyimpan dan memvisualisasikan data jejaring penelitian kesehatan di Indonesia. *Graph database* akan menghubungkan *node -node* yang berhungan satu dengan yang lainnya (Wirawan, Riyanto, and Khadijah 2016). Pada penelitian ini mencoba untuk mengusulkan membentuk pemodelan basis pengetahuan menggunakan *graph database* yang dapat memanfaatkan operasi *query graph* dengan CQL (*chyper query language*). Basis pemodelan pengetahuan yang akan dibangun pada penelitian ini nantinya akan digunakan untuk melihat riset-riset kesehatan di Indonesia pada tahun tertentu, judul-judul penelitian, topik-topik penelitian, *author*, dan afiliasi. Dari basis pengetahuan dapat dianalisis lebih lanjut salah satu implementasinya dengan *topic modelling* yang digunakan untuk melihat tren suatu topik penelitian pada tahun tertentu. Metode *topic modeling* yang digunakan adalah LDA. Hasil dari *topic modelling* dilakukan dengan tujuan untuk mengidentifikasi melakukan pemodelan topik terhadap judul-judul penelitian di bidang penelitian kesehatan. Hasil penelitian ini diharapkan dapat bermanfaat dalam bidang medis khususnya dalam pengembangan penelitian kesehatan di Indonesia

1.3 Permasalahan

Pada saat ini kebutuhan akan riset, pengembangan, dan penerapan hasil penelitian di bidang penelitian kesehatan dari waktu ke waktu semakin meningkat, baik dari kalangan akademisi, praktisi, pemerintah, atau bahkan dari kalangan umum. Angka yang dirasa cukup tinggi tersebut tentunya diikuti dengan kebutuhan akan riset atau penelitian di bidang kesehatan, baik dari kalangan peneliti, tenaga kesehatan, mahasiswa, dosen, atau kalangan akademisi lainnya di perguruan tinggi. Namun, Hal ini tidak disertai dengan pertumbuhan jumlah analisis informasi atau pengetahuan yang dapat disarikan dari dokumen karya ilmiah tersebut. Sehingga dari permasalahan tersebut peneliti mengusulkan membangun pengetahuan dan menganalisis tentang pemetaan jejaring riset jurnal ilmiah khusus di bidang kesehatan di Indonesia berdasarkan judul, topik, *author*, tahun dan afiliasi sehingga mempermudah untuk mencari informasi yang terdapat didalam penelitian di bidang kesehatan.

1.4 Rumusan Masalah

Permasalahan yang akan dipecahkan pada penelitian ini yaitu sesuai dengan latar belakang yang dijelaskan di atas dapat diuraikan sebagai berikut:

1. Bagaimana cara membangun pengetahuan jejaring penelitian kesehatan indonesia berdasarkan *author*, judul, topik, afiliasi, dan tahun dengan metode *text mining* ?
2. Bagaimana cara memvisualisasikan penelitian kesehatan Indonesia berdasarkan publikasi berdasarkan *author*, judul, topik, afiliasi, tahun dengan *graph database* ?
3. Bagaimana cara menganalisis tren penelitian kesehatan indonesia berdasarkan judul penelitian?

1.5 Batasan Masalah

Berdasarkan permasalahan di latar belakang untuk lebih fokus dan terarahnya penelitian yang dilakukan dan berdasarkan rumusan masalah yang telah dipaparkan sebelumnya maka batasan dalam penelitian ini yaitu :

1. Penelitian difokuskan pada jurnal yang berkaitan dengan penelitian di bidang kesehatan Indonesia.
2. Analisis yang digunakan menggunakan *Text Mining*
3. Penelitian ini pengambilan data dari website SINTA- *Science and Technology Index* khusus di bidang kesehatan Indonesia dengan *keyword* kesehatan dan *health*.

1.6 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah diajukan, maka dapat ditentukan tujuan dari penelitian ini. Tujuan penelitian dalam penelitian ini yaitu :

1. Mampu membangun pengetahuan jejaring penelitian kesehatan Indonesia berdasarkan *author*, judul, topik, afiliasi, dan tahun dengan metode *text mining*.
2. Dapat memvisualisasikan penelitian kesehatan Indonesia berdasarkan *author*, judul, topik, afiliasi, dan tahun.
3. Dapat menganalisis tren penelitian kesehatan Indonesia berdasarkan judul penelitian kesehatan Indonesia.

1.7 Manfaat Penelitian

Berdasarkan latar belakang, rumusan masalah, batasan masalah dan tujuan dari penelitian yang telah diuraikan pada bagian sebelumnya, adapun manfaat pada penelitian yang ingin dicapai dalam penelitian ini yaitu :

1. Hasil penelitian ini diharapkan mampu membangun pengetahuan jejaring penelitian kesehatan di Indonesia sehingga peneliti dapat melihat tren topik yang sedang diteliti di Indonesia.
2. Penelitian ini diharapkan mampu memberikan solusi untuk membantu peneliti untuk mempermudah peneliti dalam mempersiapkan penelitiannya khususnya bidang kesehatan di Indonesia.
3. Penelitian ini diharapkan mempermudah untuk memvisualisasikan dengan *Dashboard* sehingga dapat mempermudah untuk melihat peta jurnal yang berkaitan dengan bidang kesehatan di Indonesia.

1.8 Metodologi Penelitian

Pada metodologi penelitian ini menguraikan tahapan-tahapan yang dilakukan pada proses penelitian agar sesuai dan berjalan dengan baik sehingga akan mencapai tujuan yang diinginkan. Adapun tahap-tahapan yang dilakukan dalam penelitian ini adalah sebagai berikut:

1. Pengumpulan Data

Pada tahap ini dilakukan pengumpulan data tentang topik penelitian. Semua tahap pada proses pengumpulan data tersebut diperoleh dari Website Jurnal yang berkaitan tentang Penelitian Kesehatan di Indonesia. Website jurnal tentang kesehatan tersebut dilakukan *scraping* untuk mengambil data dari jurnal tersebut

berdasarkan judul, *author*, tahun, topik dan afiliasi. mengambil data dari jurnal tersebut yang dilakukan terfokus kepada penelitian kesehatan di Indonesia.

2. Pencarian studi pustaka

Pencarian studi pustaka dilakukan dengan tujuan untuk mengetahui gambaran dalam menyelesaikan permasalahan yang akan diteliti, serta mendapatkan dasar-dasar referensi yang kuat dalam menerapkan suatu metode yang akan digunakan dalam penelitian ini, yaitu dengan mempelajari studi literatur seperti artikel-artikel, dan jurnal-jurnal yang berhubungan dengan permasalahan yang akan dibahas.

3. Pengujian

Pada tahap pengujian ini merupakan tahapan dimana hasil pemodelan basis pengetahuan yang terbantu akan uji. Pengujian ini diperlukan untuk mengetahui bahwa pemodelan pengetahuan jejaring penelitian yang terbentuk dapat dijalankan sesuai dengan tujuan. Penelitian ini diharapkan mampu memberikan solusi untuk membantu peneliti untuk mempermudah peneliti dalam mempersiapkan penelitiannya khususnya bidang kesehatan di Indonesia.

4. Analisis data

Analisis ini dilakukan terhadap teks yang tidak terstruktur menjadi terstruktur dengan menggunakan teknik *tekt mining* yang diperlukan agar dapat dianalisis sesuai harapan yang dimodelkan ke dalam *graph* sehingga penelitian ini dapat divisualisasikan dan dipetakan berdasarkan penelitian di bidang kesehatan di Indonesia.

1.9 Sistematika Penulisan

Untuk mempermudah proses pembahasan dalam penelitian yang dibuat, maka dibuat sistematika penulisan pada penelitian ini :

BAB 1 Pendahuluan

Pendahuluan, merupakan pengantar terhadap permasalahan yang akan dibahas pada penelitian. Pada pendahuluan ini menguraikan tentang gambaran suatu penelitian yang terdiri dari latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metodologi penelitian, serta sistematika penulisan.

BAB 2 Landasan Teori

Bab 2 menjelaskan teori-teori yang digunakan untuk mendukung dalam memecahkan masalah pada penelitian Analisis Jejaring Penelitian Kesehatan Indonesia Menggunakan *Text Mining*. Teori-teori yang berkaitan dengan penelitian diantaranya yaitu *graph database*, *teori graph*, *text mining*, *network*, penelitian kesehatan, Jurnal SINTA, *scraping*, dan pemodelan topik.

BAB 3 Metodologi Penelitian

Bab 3 membahas tentang langkah-langkah penelitian dan gambaran umum langkah penyelesaian yang dibuat secara sistematis dan dapat dijadikan pedoman yang jelas dalam menyelesaikan permasalahan.

BAB 4 Hasil Dan Pembahasan

Bab 4 berisi tentang hasil dan pembahasan penelitian, yaitu pembahasan membentuk pemodelan basis pengetahuan dengan menggunakan *graph*, penggunaan metode *text mining* yang digunakan untuk menganalisis jejaring penelitian dibidang kesehatan di Indonesia. Hasil dari penelitian ini yaitu menganalisis jejaring penelitian kesehatan di Indonesia berdasarkan *keyword*, afiliasi, topik, tahun, judul, dan *author*.

BAB 5 Kesimpulan Dan Saran

Bab 5 menguraikan tentang kesimpulan dari seluruh bab-bab yang telah dibahas menjawab rumusan masalah dan yang perlu diperhatikan yaitu berdasar keterbatasan yang ditemukan dari penelitian dan asumsi-asumsi yang dibuat selama melakukan penelitian dan juga rekomendasi yang diuraikan untuk pengembangan penelitian selanjutnya.

BAB 2

Landasan Teori

2.1 Penelitian Sebelumnya

Semakin pesatnya perkembangan teknologi di Indonesia berpengaruh pada banyaknya penelitian yang dilakukan oleh peneliti termasuk dalam bidang kesehatan (Ministry of Health Republic of Indonesia 2016). Semakin banyaknya penelitian semakin banyak pula penyimpanan media digital yang mendorong terjadinya banyak jumlah dokumen elektronik yang tersimpan dalam repositori jurnal atau publikasi ilmiah. Berbagai karya publikasi ilmiah dari sivitas akademika seperti laporan kerja praktek, skripsi, tesis, disertasi, laporan penelitian, dan lain sebagainya telah tersedia dalam versi digital. Namun, pada umumnya hal ini tidak disertai dengan pertumbuhan jumlah analisis informasi atau pengetahuan yang dapat disarikan dari dokumen karya ilmiah tersebut. Permasalahan tersebut dapat diatasi dengan adanya *networking*/jejaring analisis yang digunakan untuk menganalisis mengetahui semua hubungan antara masing-masing pasangan publikasi dalam jaringan, pola yang dihasilkan dapat divisualisasikan dalam bentuk *graph*, menghubungkan satu aktor ke aktor lainnya relatif bersifat permanen, Pola yang terbentuk didasari oleh aturan yang mengatur keterhubungan masing-masing aktor, dapat memvisualisasikan karakteristik struktural atau hierarki dalam jejaring/*network* (Sutrisno 2017). Oleh karena itu peneliti tertarik untuk meneliti penerapan pemodelan jejaring menggunakan *text mining* dengan menggunakan *graph database neo4j* dan Metode LDA (*Latent Dirichlet Allocation*) yang digunakan untuk memetakan penelitian kesehatan di Indonesia.

Penelitian yang dilakukan (Banazir and Philip 2013) yaitu bagaimana mencari tren penelitian yang didapatkan dari dokumen abstrak jurnal dari abstrak tersebut dapat melihat tren topik apa saja yang sedang dibicarakan dalam penelitian. Selanjutnya penelitian yang dilakukan (Sri Arini, Putu Widja, and Yasa Negara 2019) menganalisis frekuensi kata yang muncul untuk mengekstrak kata kunci dari artikel ilmiah berbahasa indonesia dengan menggunakan metode *text mining*. Selanjutnya penelitian yang dilakukan (Yudiarta, Sudarma, and Ariastina 2018) berhasil mengklaster topik untuk pengelompokan berita pada data yang tidak terstruktur. Penelitian yang dilakuan (Prihatini et al. 2017) yaitu tentang mengetraksi *feature text* dengan menggunakan metode LDA hasilnya dapat mengetahui topik berita di Indonesia. Untuk menerapkan *graph* data base jejaring penelitian terkait (Wirawan and Riyanto 2017) memetakan dan untuk menyajikan

informasi *rute Bus Rapid Transit* sekaligus menyampaikan informasi perpindahan koridor di Semarang dengan *graph database* sehingga dapat dilihat rute jejaringnya. Penelitian selanjutnya (Bajaj et al. 2018) yaitu bagaimana memetakan rekomendasi untuk resep makanan sehingga mudah untuk melihat rekomendasi yang diterapkan menggunakan *graph database*. Penelitian yang dilakukan (Constantinov et al. 2018) bagaimana melihat tren tenaga kerja dan informasi perusahaan yang sedang dicari oleh calon tenaga kerja.

Berdasarkan penelitian dan permasalahan-permasalahan tersebut metode *text mining* dan *graph database* ini merupakan metode yang cocok dalam pengembangan tersutrukturnya data informasi yang dapat diterapkan untuk mengatasi masalah mengetahui tren topik (Jo 2018). Algoritma-algoritma dalam *text mining* dibuat untuk dapat mengenali data yang sifatnya terstruktur. Dalam penelitian ini metode *text mining* digunakan untuk mengetahui topik-topik penelitian kesehatan di Indonesia. Metode *text mining* yang digunakan dalam penelitian ini adalah *topic modelling LDA (Latent Dirichlet Allocation)*. Metode LDA digunakan untuk mengetahui tren topik penelitian kesehatan di Indonesia. Metode LDA digunakan cocok untuk pemodelan topik yang menerapkan pengelompokan variabel jumlah teks yang besar (Lakshmi Prasanna and Rajeswara Rao 2019). Sumber data diperoleh dari website jurnal SINTA yang berkaitan dengan penelitian kesehatan di Indonesia. Cara untuk memperoleh data dengan teknik *web scraping* dari dokumen-dokumen yang terdapat di jurnal/website publikasi ilmiah SINTA. Selain itu juga diterapkan model *network* analisis untuk mempelajari hubungan antar satu unit entitas dengan unit entitas lainnya dengan bantuan teori *graph* (Wirawan et al. 2019). Penerapan pemodelan jejaring menggunakan *graph database* yang digunakan untuk memodelkan jejaring berdasarkan data set penelitian kesehatan di Indonesia dengan menggunakan *tools graph database Neo4j*.

Berdasarkan penelitian-penelitian di atas maka dapat disimpulkan bahwa penelitian belum fokus dapat memetakan jurnal berdasarkan topik dan afiliasi jurnal publikasi yang berkaitan dengan kesehatan di Indonesia. Oleh karena itu, penelitian yang akan dibahas oleh penulis yaitu berjudul “*Analisis Jejaring Penelitian Kesehatan Indonesia Menggunakan Text Mining*”. Penelitian ini nantinya diharapkan mampu memetakan jurnal penelitian dibidang kesehatan di Indonesia sehingga peneliti dapat melihat trend topik yang sedang diteliti di Indonesia. Penelitian ini juga diharapkan mampu memberikan solusi untuk membantu peneliti untuk mempermudah peneliti dalam mempersiapkan penelitiannya khususnya bidang kesehatan di Indonesia. Penelitian selanjutnya diharapkan mempermudah untuk memvisualisasikan dengan *Dashboard* sehingga dapat

mempermudah untuk melihat peta jurnal yang berkaitan dengan bidang kesehatan di Indonesia. Adapun penelitian - penelitian sebelumnya yang pernah dilakukan berkaitan dengan penggunaan metode *text mining* disajikan pada Tabel 2.1.

Tabel 2.1 Penelitian Terdahulu

Peneliti	Judul	Metode	Data	Hasil
Ni Wayan Sri Arini, Ida Bagus Putu Widja, Komang Rinarta (2019)	“Analisis Frekuensi Kata untuk Mengekstrak Kata Kunci dari Artikel Ilmiah Berbahasa Indonesia”	Algoritma TF (<i>Term Frequency</i>) - IDF ()	Artikel Ilmiah Berbahasa Indonesia	Kata kunci yang dihasilkan dari artikel diambil hanya 20 kata
(Budhi Kurniawan Wangsa, Darmawan Utomo, Saptadi Nugroho, 2014)	“Sistem Peringkat Berita Otomatis berbasis <i>Text Mining</i> menggunakan GVSM”	GVSM (<i>Generalized Vector Space Model</i>)	Media Massa Online	Berita yang ringkas tidak bisa langsung dinilai baik oleh responden keberhasilan hanya mencapai 53%.
(Syukriyanto Latif, 2018)	<i>Text Mining For Classification Of Content Abstract Journal English Using Dimensional Reduction Method And Naive Bayes</i>	Metode Reduksi Dimensi Dan Naive Bayes	Konten Abstrak Jurnal Bahasa Inggris	Akurasi 95,33%, dengan skenario 75% data latihan - 25% data testing
(Nyoman Gede Yudiarta, Made Sudarma, Wayan Gede)	Penerapan Metode <i>Clustering Text Mining</i> Untuk Pengelompokan	<i>Clustering Text Mining</i>	Berita	pengujian 50 data memiliki rata – rata persentase nilai Precision dan

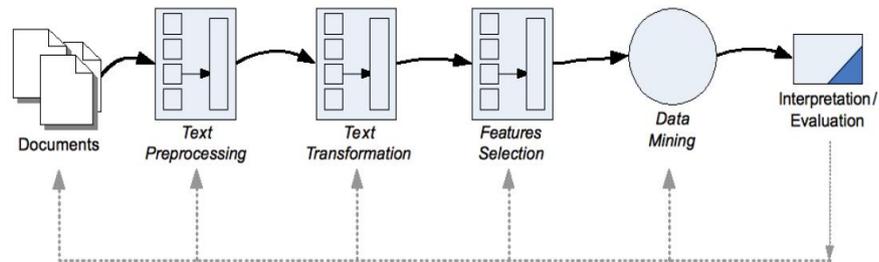
Ariastina, 2018)	Berita Pada <i>Unstructured Textual Data</i>			Recall yang paling besar yaitu 76,92% untuk precision dan sebesar 79,58% untuk recall nya.
(Yoga Sahria, 2019) Peneliti Yang Akan Dikembangkan	<i>Indonesian Medical Research Network Analysis Through Text Mining</i>	<i>Graph Database dan Text Mining</i>	Jurnal/ publikasi Kesehatan di Indonesia pada SINTA	Dapat Memetakan, memvisualisasikan, Menganalisis Jurnal di Bidang Kesehatan Indonesia.

2.2 Text Mining

Text mining adalah sebuah proses analisis dalam data yang berupa teks dimana sumber data didapatkan dari dokumen (Feldman and Sanger 2006). *Text Mining* merupakan ilmu yang sedang hangat, yang sebelumnya tidak dikenal dalam informasi, *text mining* secara otomatis dapat mengekstraksi informasi dari berbagai tulisan sumber daya. Elemen kunci adalah menghubungkan bersama dari yang diekstraksi informasi bersama untuk membentuk fakta baru atau hipotesis baru untuk dieksplorasi selanjutnya dengan cara eksperimen yang lebih konvensional. *Text mining* umumnya digunakan dalam klasifikasi dokumen tekstual yang mana dokumen-dokumen tersebut akan diklasifikasikan sesuai dengan topik dokumen tersebut. Metode *text mining* dapat membantu suatu artikel dapat diketahui jenis kategorinya melalui kata-kata yang terdapat pada artikel tersebut. Kata-kata yang dapat mewakili isi dari artikel tersebut dicocokkan dan dianalisa pada basis data kata kunci yang telah ditentukan sebelumnya. *Text mining* bertujuan untuk penganalisisan teks guna menyorikan informasi yang bermanfaat untuk tujuan tertentu. Sehingga metode *text mining* dapat membantu melakukan pengelompokan suatu dokumen dengan waktu yang singkat. Tahapan *text mining* dalam melakukan analisa yaitu dengan cara melakukan pengumpulan data kemudian melakukan ekstraksi terhadap fitur yang akan digunakan.

Proses preprosesing metode *text mining* diawali dengan ketidakaturan struktur data teks, maka proses *text mining* memerlukan beberapa tahap awal yaitu untuk

mempersiapkan agar teks dapat diubah menjadi lebih terstruktur. *Text mining* sebagai proses secara otomatis untuk mengekstrak informasi yang bermakna, berguna, dimana sebelumnya tidak diketahui dan pada akhirnya dapat di pahami dari penyimpanan dokumen tekstual. Proses *text mining* dapat dilihat pada Gambar 2.1.



Gambar 2.1. Proses *Text Mining*

Pada Gambar 2.1 Tahap pertama dalam proses *Text Mining* klasifikasi dokumen atau kategorisasi dokumen. Klasifikasi dokumen ini digunakan dalam persoalan menetapkan dokumen satu atau lebih untuk dikategorikan. Tahap kedua yaitu *Text Preprocessing* adalah tahapan dimana melakukan seleksi data yang yang tidak terstruktur akan diproses pada setiap dokumen. *Text Preprocessing* pemrosesan awal teks ini bertujuan untuk mempersiapkan teks sebelum diolah sebelum lanjut. Proses tahapan *preprocessing* yaitu meliputi *tokenizing*, *case folding*, *filtering*, dan *stemming*. Pada tahap ketiga *Text Transformation* yaitu tahapan yang dipergunakan untuk mengubah kata-kata ke dalam bentuk dasar, sekaligus untuk mengurangi jumlah kata- kata tersebut. Proses pendekatan yang dapat dilakukan yaitu dengan *stemming* dan *stopwords* penghapusan. Pada tahap empat yaitu *Feature Selection* adalah sebuah cara yang digunakan untuk dapat mengoptimalkan kinerja dari *classifier*. Pada *feature selection* cara kerjanya pengurangan ruang fitur yang besar, yaitu mengeliminasi atribut yang kurang tepat atau kurang relevan serta dengan menggunakan penggunaan algoritme *feature selection* yang tepat sehingga dapat meningkatkan akurasi. Tahap kelima *Process Data Mining* bertujuan untuk menghasilkan patterns yang berguna dari koleksi *text*. Aktivitas *text mining* untuk *step* data mining terdiri dari pemilihan mining teknik yang benar, penentuan mining model dan parameters. Tahap terakhir adalah *Interpretation/Evaluation* tujuannya sebagai *text Summarization* yaitu *summarization* untuk mengidentifikasi *key content* dari berbagai sumber data dimana *key content* ini bisa merepresentasikan keseluruhan *text* secara akurat. *Text summarization* bisa juga disamakan dengan *text classification text summarization* digunakan untuk menjelaskan seluruh konten *text* dengan mengekstrak hanya *keyword* yang penting.

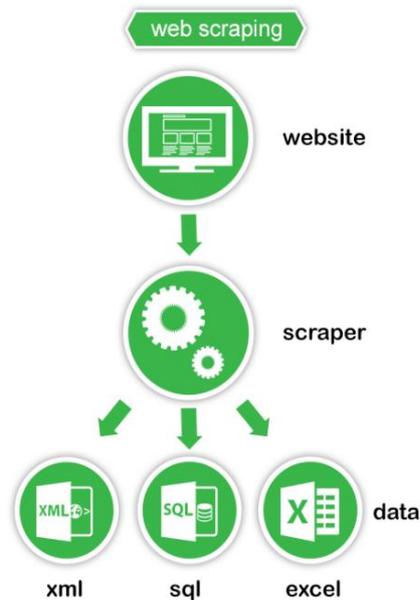
Text mining adalah penambahan dari sebuah data yang dilakukan oleh komputer untuk mendapatkan suatu hal baru yang sebelumnya belum pernah diketahui dan menemukan informasi yang tersirat secara implisit yang berasal dari data yang diekstrak secara otomatis yang berasal dari sumber yang berbeda (feldman dan sanger 2006). *Text mining* ini merupakan cabang dari metode *Data Mining*. Oleh karena itu, metode *Text Mining* hampir mirip dengan penelitian data *mining*. Perbedaan antara *text mining* dan data *mining* adalah pola yang digunakan, pada *Text Mining* data yang diambil adalah sekumpulan bahasa alami *Natural Language* yang tidak terstruktur sedangkan pada data *mining* pola yang diambil adalah dari kumpulan data yang notabene sudah terstruktur dengan baik (han dan kamber, 2006).

Text mining secara umum atau luas dapat didefinisikan sebagai proses pengetahuan intensif yang mana pengguna dapat berinteraksi dengan koleksi dokumen dari waktu ke waktu dengan menggunakan seperangkat alat analisis. Metode *Text mining* ini mampu mengekstrak informasi yang dapat dianalisis dan berguna dari sumber data melalui identifikasi dan eksplorasi pola yang menarik. Metode penelitian *text mining* hampir mirip mengarah pada bidang penelitian data mining. Oleh karena itu, bahwa metode *text mining* dan *data mining* berada pada tingkat arsitektur yang hampir sama. Penambangan teks atau *text mining* dapat terbagi menjadi dua proses tahap pertama diawali dengan menerapkan struktur data terhadap sumber data teks dan dilanjutkan dengan proses ekstraksi informasi dan pengetahuan yang relevan dengan sumber data teks terstruktur dengan menggunakan teknik dan alat yang sama dengan penambangan data. Dari beberapa penulis yang mendefinisikan *text mining* dapat disimpulkan bahwa *text mining* merupakan kegiatan atau proses penambahan teks dan menganalisis data dalam jumlah besar secara otomatis kemudian diklasifikasikan. *Text mining* sering digunakan untuk menganalisis informasi, pengambilan keputusan, dan tugas-tugas manajemen informasi lainnya. Saat ini *text mining* telah banyak digunakan untuk berbagai bidang contohnya dibidang biomedis, pengembangan software, pengembangan aplikasi, media *online*, akademik, pemasaran, dan keamanan.

2.3 Web Scraping

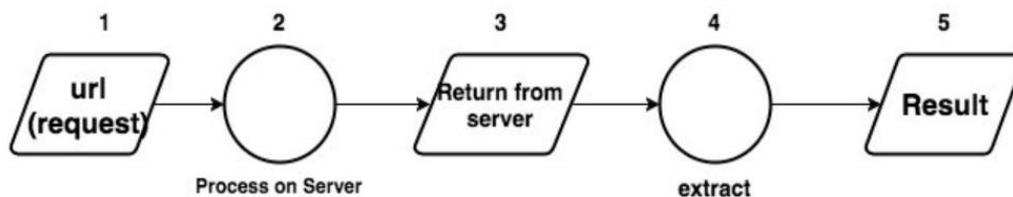
Web Scraping adalah metode pengumpulan data melalui internet, meskipun *web scraping* bukan suatu hal yang baru, belakangan ini metode *web scraping* sangat populer digunakan untuk pemenuhan *data mining*. Sebelumnya metode ini dikenal dalam beberapa istilah, diantaranya *screen scraping*, *data mining*, *web harvesting* ataupun metode lain yang sejenis. Menurut teori, *web scraping* adalah cara untuk mengumpulkan data

menggunakan metode yang berbeda dengan penggunaan API (*Application Programming Interface*). Cara seperti ini biasanya dimulai dengan penulisan kode program yang mana biasanya digunakan sebagai otomatisasi *query* untuk melakukan *request* data terhadap *server*, umumnya data tersebut dapat berbentuk *form html* dan *file* lain yang berhubungan *web pages*. Data hasil *request* tersebut dapat dilakukan ekstraksi untuk menghasilkan informasi yang akan dicari (mitchell, 2015). Berikut langkah *web scraping* dapat dilihat pada Gambar 2.2.



Gambar 2.2 Langkah *Web Scraping*

Web scraping adalah proses meminta dokumen web secara otomatis dan mengumpulkan informasi darinya. Sebenarnya, untuk melakukan *scraping* web, harus melakukan beberapa tingkat mengakses web untuk *request* pada situs web. Adapun proses *web scraping* dapat dilihat pada Gambar 2.3.



Gambar 2.3 Proses *Scraping*

Pada Gambar 2.3 Proses pertama *Request* url yang dijadikan target, proses kedua *Request* diproses oleh server target. Proses ketiga hasil dari *request* dari url (hasilnya adalah teks dengan format HTML), Kemudian proses selanjutnya mengekstrak data

(mengambil teks yang diperlukan dari tahap ke-3), Hasil yang ekstrak (menentukan *output* yang diinginkan).

2.4 Jurnal SINTA

Sinta (*Science and Technology Index*) diinisiasi pada tahun 2016 oleh Direktorat dari Jenderal Penguatan Penelitian dan Pengembangan, Kementerian Riset Teknologi dan Pendidikan Tinggi Republik Indonesia, dengan melibatkan para ahli dari berbagai lembaga. Konten Sinta dari jurnal Indonesia yang telah diterbitkan secara elektronik memiliki profil atau *google scholar*, *preview scopus* dan berisi sejumlah kutipan, indeks h, indeks i-10, pengembangan Sinta selanjutnya akan mencakup makalah, buku dan paten peneliti di Indonesia, dan profil penulis dari *google scholar*. Sinta menyediakan perbaruan data penulis, lembaga, dan penerbit jurnal yang sejak tahun 2017.

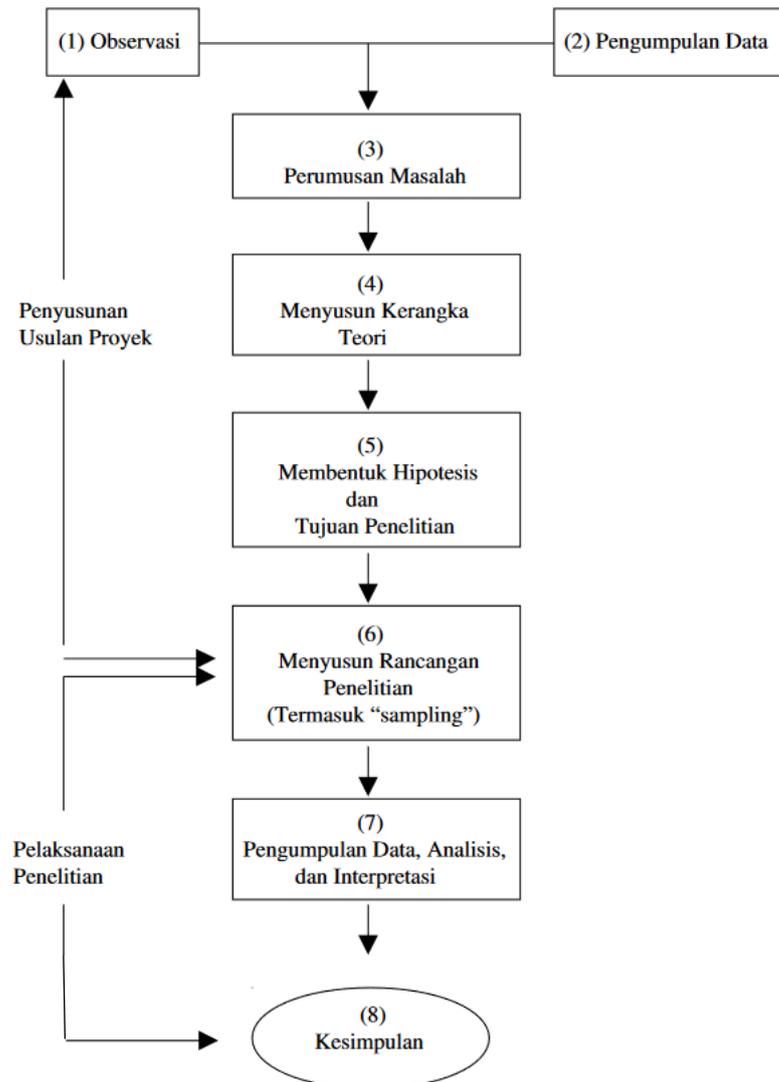
Sinta (*Science and Technology Index*) yaitu tempat penyimpanan jurnal yang dibuat untuk menyimpan hasil penelitian-penelitian di Indonesia, salah satunya adalah penelitian kesehatan di Indonesia. Jurnal SINTA Sistem informasi penelitian berbasis web menawarkan akses cepat, mudah dan komprehensif untuk mengukur kinerja para peneliti, lembaga dan jurnal di Indonesia. SINTA ini dapat dijadikan tolok ukur dan analisis, identifikasi kekuatan penelitian masing-masing lembaga untuk mengembangkan kemitraan kolaboratif, untuk menganalisis tren penelitian dan direktori ahli. Visi dari SINTA adalah “Menjadi referensi kinerja penelitian”. Misi SINTA mengembangkan Citasi/kutipan dan keahlian di Indonesia, Mengidentifikasi dan analisis kekuatan penelitian masing-masing institusi, Mengembangkan sistem analisis tren penelitian di Indonesia, Mengembangkan direktori keahlian di Indonesia.

2.5 Penelitian Kesehatan

Menurut (Dharminto 2018) Penelitian (*research*) merupakan rangkaian kegiatan ilmiah dalam rangka pemecahan suatu permasalahan. Jadi sebuah penelitian adalah bagian dari sebuah usaha pemecahan suatu masalah. Adapun fungsi penelitian adalah mencari sebuah penjelasan dan jawaban terhadap permasalahan serta memberikan alternatif bagi kemungkinan yang dapat digunakan untuk pemecahan masalah. Penelitian suatu proses penyelidikan yang dilakukan secara aktif, tekun, dan sistematis, dimana tujuannya untuk menemukan, menginterpretasikan, dan merevisi fakta-fakta.

Penelitian merupakan suatu proses investigasi secara sistematis dengan cara mempelajari berbagai bahan dan sumber untuk membangun fakta-fakta dan mencapai kesimpulan baru. Adapun Tujuan penelitian merupakan suatu proses mendapatkan atau

menemukan suatu data untuk keperluan dan tujuan tertentu. Secara umum tahapan penelitian disajikan pada Gambar 2.4.



Gambar 2.4 Proses Penelitian

Secara umum tahapan-tahapan penelitian pada umumnya terdapat tujuh tahap satu sama lain yang bergantung dan berhubungan. Dengan kata lain masing-masing tahap saling mempengaruhi dan dipengaruhi tahap-tahap berikutnya. Adapun tujuh tahap penelitian yaitu perencanaan, pengkajian secara teliti terhadap rencana penelitian, pengambilan sampling, penyusunan daftar pertanyaan, kerja lapangan, editing dan coding, analisis dan laporan.

Menurut (Wardhani and Paramita 2016) Kesehatan merupakan keadaan sehat sejahtera dari jiwa, badan dan sosial yang dapat memungkinkan setiap orang hidup produktif secara sosial dan ekonomis. Pengertian sehat menurut (WHO) Organisasi Kesehatan Dunia pada tahun 1948 yaitu “suatu keadaan mental, fisik dan sosial

kesejahteraan dan bukan hanya ketiadaan dari sebuah penyakit atau kelemahan". Kata dasar dari kesehatan berasal dari kata sehat yang artinya terbebas dari segala semua gangguan atau penyakit baik penyakit fisik maupun psikis yang dialami. Jika diartikan dari kata dasarnya, maka kesehatan adalah sebuah kondisi atau sebuah keadaan yang menggambarkan tubuh terbebas dari segala penyakit atau gangguan fisik dan psikis. Kesehatan juga dapat diartikan keadaan yang tenang tentram dan sejahtera dari badan, jiwa, dan sosial yang memungkinkan setiap orang hidup secara produktif dalam bidang sosial, dan ekonomis.

Berdasarkan Undang-Undang Nomor 12 Tahun 2012 tentang Pendidikan Tinggi, penelitian merupakan kegiatan yang dilakukan menurut sebuah metode dan kaidah ilmiah secara sistematis untuk kemudahan diperoleh informasi, data, dan keterangan berkaitan dengan pengujian dan/atau pemahaman dari cabang ilmu pengetahuan dan teknologi.

Kesehatan menurut Kementerian Kesehatan Indonesia adalah keadaan sejahtera dari badan, jiwa, dan sosial yang memungkinkan setiap orang hidup produktif secara sosial, dan ekonomis. Upaya kesehatan adalah setiap kegiatan untuk memelihara, dan meningkatkan kesehatan yang dilakukan oleh pemerintah, dan atau masyarakat. Tenaga kesehatan adalah setiap orang yang mengabdikan diri dalam bidang kesehatan serta memiliki pengetahuan, dan atau keterampilan melalui pendidikan di bidang kesehatan yang untuk jenis tertentu memerlukan kewenangan untuk melakukan upaya kesehatan. Sarana kesehatan adalah tempat yang digunakan untuk menyelenggarakan upaya kesehatan. Kesehatan adalah sesuatu yang sangat berguna bagi kesehatan seseorang untuk menjalankan aktifitas sehari-hari.

Publikasi ilmiah merupakan sistem publikasi yang dilakukan melalui proses penelitian atau pemeriksaan terhadap suatu karya ilmiah oleh pakar lain dalam bidang yang sama. Menurut KBBI, publikasi bermakna pengumuman atau penerbitan. Publikasi ialah hasil tulisan atau karya tulis yang diterbitkan atau dipublikasikan. Penerbitan karya tulis dapat bersifat populer, populer-ilmiah, dan ilmiah (Nashihuddin, 2016).

Publikasi ilmiah adalah sistem publikasi yang dilakukan berdasarkan peer review dalam rangka untuk mencapai tingkat obyektivitas setinggi mungkin. "Sistem" ini, bervariasi tergantung bidang masing-masing, dan selalu berubah, meskipun seringkali secara perlahan. Sebagian besar karya akademis diterbitkan dalam jurnal ilmiah atau dalam bentuk buku. Sebagian besar bidang akademik yang telah mapan memiliki jurnal dan bentuk publikasi tersendiri, meskipun banyak pula terdapat jurnal akademik yang bersifat interdisipliner (antar cabang) dan mempublikasikan karya dari beberapa bidang yang

berbeda. Jenis-jenis publikasi yang dapat diterima sebagai kontribusi terhadap bidang ilmu pengetahuan dan penelitian sangat bervariasi di antara berbagai bidang.

Penelitian Kesehatan adalah langkah metode ilmiah yang berorientasikan atau memfokuskan kegiatannya pada masalah-masalah yang timbul di bidang kesehatan. Kesehatan itu sendiri terdiri dari dua sub bidang pokok, yakni pertama kesehatan individu yang berorientasikan klinis dan pengobatan. Sub bidang kedua yang berorientasi pada kelompok atau masyarakat, yang bersifat pencegahan. Selanjutnya sub bidang kesehatan ini pun terdiri dari berbagai disiplin ilmu, seperti kedokteran, keperawatan, epidemiologi, pendidikan kesehatan, kesehatan lingkungan, manajemen pelayanan kesehatan, gizi dsb. Sub bidang tersebut saling berkaitan dan mempengaruhi kondisi kesehatan masyarakat pada umumnya.

Tujuan penelitian kesehatan menurut (Notoatmodjo, 1993) menemukan atau menguji fakta baru maupun fakta lama yang berhubungan dengan bidang kesehatan, Melakukan analisis terhadap hubungan antara fakta-fakta yang ditemukan dalam bidang kesehatan, menjelaskan tentang fakta yang ditemukan serta hubungannya dengan teori yang telah ada, mengembangkan metode atau konsep baru dalam pelayanan kesehatan untuk meningkatkan kesehatan masyarakat. Manfaat dari penelitian kesehatan Hasil penelitian dapat digunakan untuk menggambarkan tentang keadaan atau status kesehatan individu, kelompok atau masyarakat, menggambarkan kemampuan sumber daya dan kemungkinan sumber daya tersebut guna mendukung pengembangan pelayanan kesehatan, Dapat dijadikan bahan kajian untuk mencari sebab masalah kesehatan atau kegagalan yang terjadi dalam pelayanan kesehatan. Sehingga dapat dijadikan acuan untuk mencari solusi atau alternatif penyelesaian masalah, serta dapat dijadikan sebagai sarana untuk menyusun kebijakan pengembangan pelayanan kesehatan.

2.6 Metode Social Network

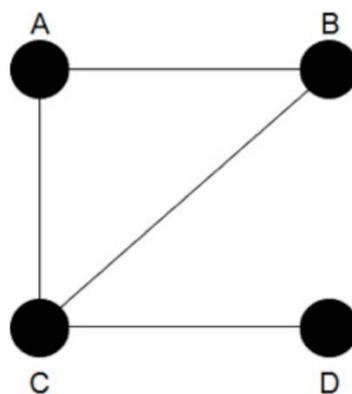
Social Network adalah salah satu metode *Social Computing* untuk melakukan ekstrak informasi pada data yang tidak terstruktur dan memiliki volume yang besar. *Social network* merupakan pemodelan terhadap user yang dilambangkan dengan titik (*nodes*) dan interaksi antar user tersebut yang dilambangkan dengan garis (*edges*) (Bratawisnu and Alamsyah 2018). *Social network* memiliki beberapa properti jaringan untuk memetakan hubungan yang sangat membantu untuk meningkatkan penciptaan *management knowledge* di organisasi. *Social Network* mempelajari struktur hubungan yang mengaitkan individu atau unit sosial lain serta ketergantungan dalam perilaku atau sikap yang berhubungan dengan susunan hubungan sosial. Hubungan tersebut digambarkan dengan *nodes* dan *ties* (atau

disebut juga *edges*, *links* atau *connections*). *Node* merupakan aktor dalam suatu jaringan, sedangkan *ties* adalah garis yang menghubungkan satu *node* dengan yang lain (O'Malley and Marsden 2008). *Social Network* adalah suatu proses yang menelaah struktur sosial melalui penggunaan *Network Science*. *Network Science* merupakan bidang kajian akademik baru yang mempelajari jaringan kompleks seperti jaringan telekomunikasi, jaringan komputer, jaringan biologis, dan jaringan sosial. Kajian ini mendasarkan dirinya pada *Graph Theory* (Yusainy, Chawa, and Kholifah 2017).

2.7 Teori Graph

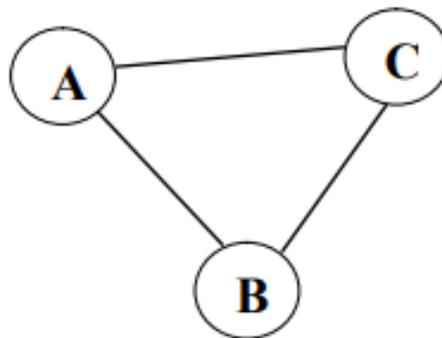
Teori *graph* atau graf pertama kali dicetuskan oleh Leonhard Euler ahli dibidang matematika yang berasal dari Swiss. Ide awal pertama yaitu mencari jalur paling baik untuk melintasi jembatan Koningberg, Prussia. Beliau menggambarkan ada beberapa jalur untuk melintasi ketujuh jembatan yang terdapat di Koningberg. Jalur yang melintasi digambarkan menuju titik-titik tujuan dari ketujuh jembatan tersebut dengan sekali melintasi setiap jembatan. Setiap jalur tersebut dikenal dengan *Eulerian Path*.

Teori *Graph* merupakan cabang ilmu matematika diskrit yang banyak penerapannya dalam berbagai bidang ilmu seperti *engineering*, fisika, biologi, kimia, arsitektur, transportasi, teknologi komputer, ekonomi, sosial dan bidang lainnya. *Graph* adalah sebuah objek dasar pelajaran dalam teori graf. Dalam kehidupan sehari-hari, sebuah graf merupakan himpunan dari objek-objek yang dinamakan *titik*, *simpul*, atau *sudut* dihubungkan oleh penghubung yang dinamakan *garis* atau *sisi* (Rabuzin and Sestak 2018). Dalam graf yang memenuhi syarat biasanya *tidak berarah*, sebuah garis dari titik *A* ke titik *B* dianggap sama dengan garis dari titik *B* ke titik *A*. Dalam *graf berarah*, garis tersebut memiliki arah. *Graph* pada dasarnya digambarkan dengan bentuk diagram sebagai himpunan dari titik-titik (sudut atau simpul) yang digabungkan dengan kurva (garis atau sisi). Adapun menurut Rabuzin contoh *graph* disajikan pada Gambar 2.4.



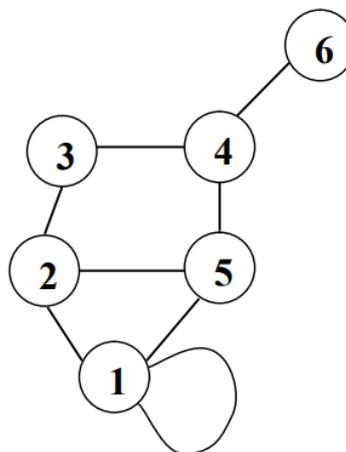
Gambar 2.4 Contoh *Graph*

Teori *Graph* merupakan cabang ilmu matematika diskrit yang banyak penerapannya dalam berbagai bidang ilmu seperti *engineering*, fisika, biologi, kimia, arsitektur, transportasi, teknologi komputer, ekonomi, sosial dan bidang lainnya. *Graph* adalah himpunan pasangan tak berurut antara *vertex* (titik atau *node*) dan *edge* (garis atau *arcs*). *Graph* merupakan sebuah metode untuk mencari solusi dari masalah ini adalah dengan membentuk model yang disebut dengan *edge*. Dalam teori *graph*, *adjacency list* merupakan bentuk representasi dari seluruh sisi atau busur dalam suatu *graph* sebagai suatu senarai. Simpul-simpul yang dihubungkan sisi atau busur tersebut dinyatakan sebagai simpul yang saling terkait. Dalam implementasinya, *hash table* digunakan untuk menghubungkan sebuah simpul dengan senarai berisi simpul-simpul yang saling terkait tersebut. Adapun gambar *undirected graph* disajikan pada Gambar 2.5.



Gambar 2.5 *Undirected Graph*

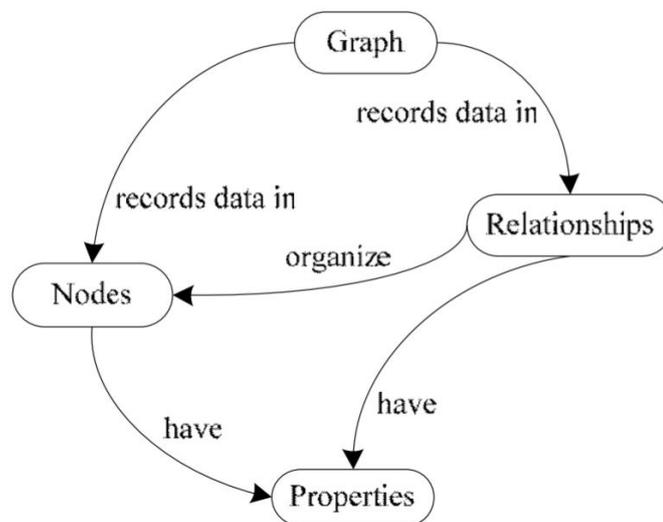
Adjacency Matrix merupakan representasi matrik yang berhubungan antar simpul *graph*. Baris dan kolom dari matrik tersebut mempresentasikan simpul-simpul dan nilai dalam metrik yang menyatakan hubungan antar simpul. Gambar *graph* adjancency matrik dapat dilihata pada Gambar 2.6.



Gambar 2.6 *Graph* tak berarah berlabel

Dalam bidang ilmu komputer, sebuah *graph* dapat dinyatakan sebagai sebuah struktur data, atau secara spesifik dinamakan sebagai ADT (*abstract data type*) yang terdiri dari kumpulan simpul dan sisi yang membangun hubungan antar simpul. Konsep ADT *graph* ini merupakan turunan konsep *graph* dari bidang kajian matematika. Teori *graph* adalah merupakan pecahan cabang dari bidang matematika diskrit yang mempunyai cabang diberbagai bidang.

Graph database merupakan keterhubungan antar data yang optimal dan membentuk suatu karakter spesifik. *Graph database* dimodelkan dengan *nodes* dan *property* menghubungkan antar data menggunakan *edges* (Wita, Bubphachuen, and Chawachat 2018). *Database Neo4j* adalah sebuah database grafik *open source* berbasis java dengan kinerja yang tinggi (Ma et al. 2014). *Neo4j* salah satu *graph database* yang dapat memvisualisasikan *node-node* yang saling terhubung yang populer dalam era komputasi saat ini (Johnpaul and Mathew 2017). *Neo4j* menyediakan cara mudah untuk memvisualisasikan data dengan struktur grafik yang baik. Pengguna dapat mendeklarasikan atribut data dalam bentuk *node* dan hubungan antara *node* dan dapat mendeklarasikan sendiri. Adapun struktur *Neo4J* menurut Yisong ditunjukkan pada Gambar 2.7.



Gambar 2.7 Struktur Database Neo4j

Dalam penelitian ini pada tahap implementasi *graph database Neo4j nodes*, properti, relasi mempresentasikan judul, topik, *author*, afiliasi, dan tahun pada penelitian kesehatan di Indonesia. Tahap berikutnya menyusun algoritma yang dapat digunakan untuk memodelkan jejaring penelitian. Tahap selanjutnya yaitu dengan menerapkan *query CQL (chyper query language)* yang *support* dengan *neo4j* yaitu *create, match, return, where,*

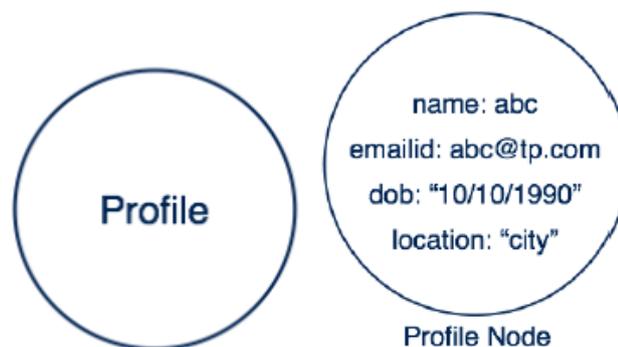
marge, unique, load, delete (Dharmawan and Sarno 2017). Penerapan CQL *query* untuk mengambil informasi berdasarkan *keyword*, judul-judul penelitian pada tahun tertentu, afiliasi-afiliasi penelitian, *author* penelitian, topik penelitian dan tahun penelitian khususnya di bidang penelitian di Indonesia.

Neo4j adalah salah satu *software graph database* dengan menerapkan visualisasi berdasarkan *graph* yang ada. Bahasa yang digunakan dalam *tools Neo4j* adalah Java dengan *Neo Technology*. *Neo4j* merupakan *software* yang disediakan berbasis *open source*, No SQL, dan merupakan *graph database* dikenal sebagai sistem manajemen grafik database. *Graph database* adalah data base atau basis data yang menyimpan sekumpulan informasi dalam bentuk grafik yang terstruktur. Penyimpanan data dalam *graph database* dinotasikan dengan sebuah *node*. Adapun perbedaan yang mendasar perbandingan *graph database* dengan *Relational Database Management System* disajikan pada Tabel 2.2.

Tabel 2.2 *Graph Database VS RDBMS*

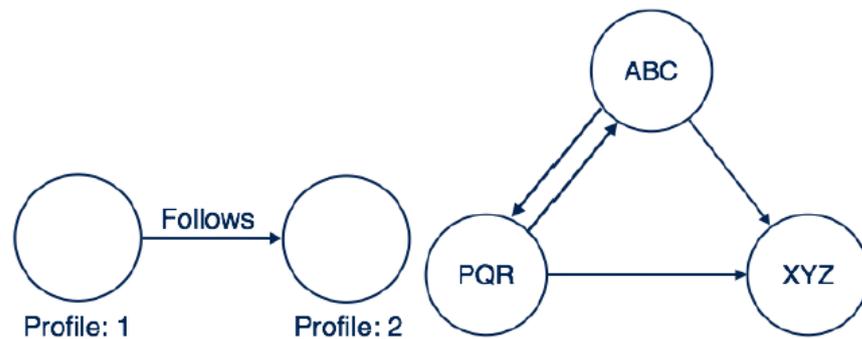
No	<i>Graph Database</i>	<i>RDBMS</i>
1	Grafik	Tabel
2	<i>Node</i>	Baris
3	Sifat dan nilai	Kolom dan Data
4	Hubungan	Kendala
5	Traversal	Bergabung

Dasar Teori *Graph* adalah sebuah set *node* dan hubungan yang menghubungkan simpul, Sebuah *node* dapat memiliki suatu hubungan dengan *node* yang lain melalui arah yang dimiliki *node* tersebut. Daintara komponen yang membentuk sebuah *graph database* adalah *node*, properti, label, hubungan dan data *browser*. *Node* sebuah grafik yang paling sederhana adalah *node* tunggal ditunjukkan pada Gambar 2.8.



Gambar 2.8 *Node dan value*

Node adalah bagian dasar fundamental dari sebuah *graph database*. *Node* pertama pada Gambar 2.8 tidak mengandung sifat, untuk menambahkan suatu penjelas dapat ditambahkan dengan beberapa properti pada *node* yang terbentuk. *Node* kedua berisi *value* suatu properti yaitu berisikan nama, email, tanggal lahir, dan lokasi. Dimana properti merupakan pasangan kunci nilai untuk merepresentasikan atau menggambarkan *node* grafik dan hubungan. Hubungan antar *node* satu arah dan lebih ditunjukkan pada Gambar 2.9.



Gambar 2.9 Hubungan antar *node* satu arah dan lebih

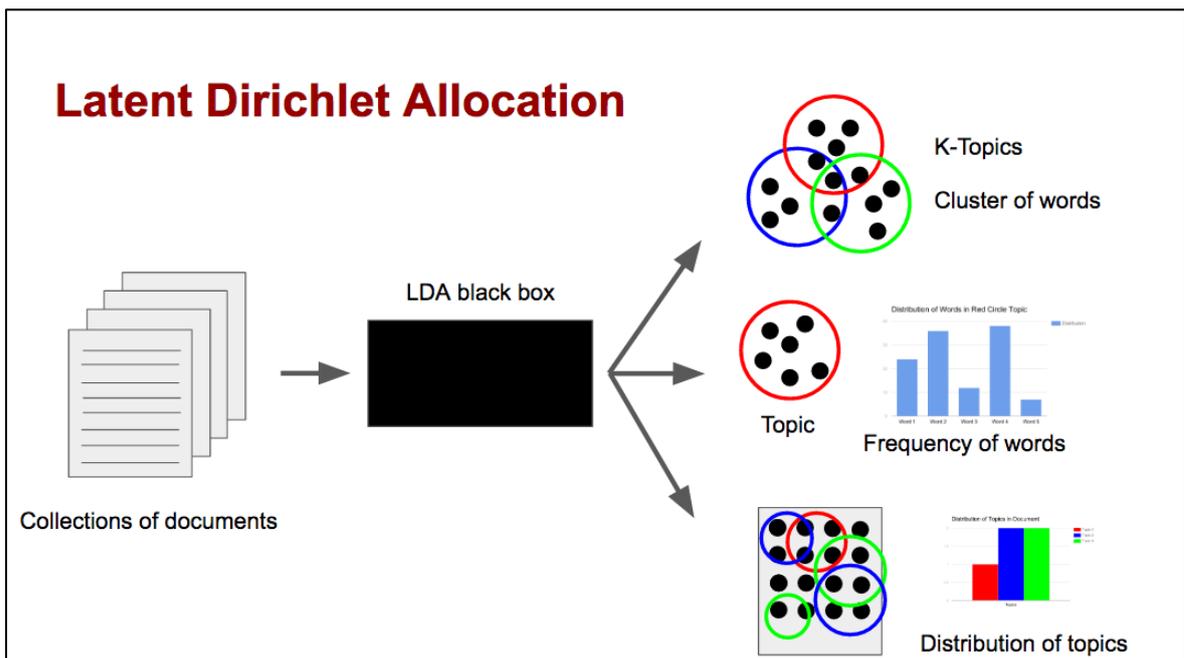
Pada gambar 2.9 menunjukkan antara dua *node* yang memiliki satu hubungan arah yang menunjukkan bahwa Profil 1 mengikuti Profil 2. Pada kasus yang kompleks seperti pada gambar 2.9 sebelah kanan dengan memiliki lebih dari satu arah sebagaimana hubungan dari XYZ ke PQR merupakan hubungan searah sedangkan hubungan ABC ke XYZ yaitu hubungan dua arah. Adapun query yang disediakan database neo4J yaitu CQL. CQL (*Chyper Query Language*) merupakan bahasa pemrograman yang digunakan oleh *Neo4j* untuk melakukan interaksi sebuah perintah yang digunakan dalam menjalankan sebuah program. Adapun Perintah yang digunakan ditunjukkan pada Tabel 2.3.

Tabel 2.3 Sintak pada CQL

No	Perintah	Penggunaan
1	CREATE	Membuat <i>node</i> , hubungan dan sifat
2	MATCH	Mengambil data tengah <i>node</i> , hubungan dan sifat
3	RETURN	Kembali ke hasil query
4	WHERE	Memberikan kondisi untuk menyaring data pengambilan
5	DELETE	Menghapus <i>node</i> dan hubungan
6	REMOVE	Menghapus sifat <i>node</i> dan hubungan
7	ORDER BY	Menyortir data pengambilan
8	SET	Menambah atau memperbarui label

2.8 Pemodelan Topik

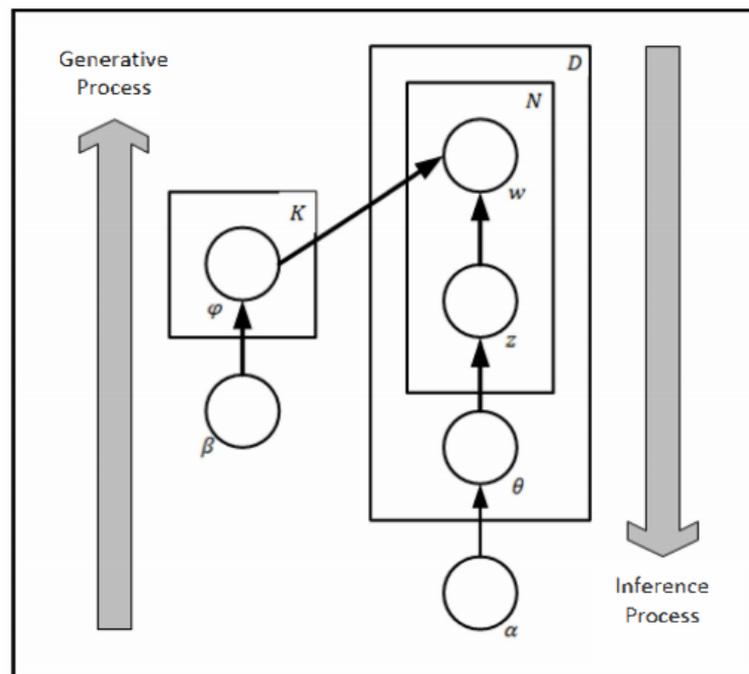
Pemodelan topik adalah metode untuk menemukan topik dan tren dari *corpus*. Pemodelan topik telah diterapkan di beberapa bidang. Metode pemodelan topik umumnya digunakan untuk secara otomatis mengatur, memahami, mencari, dan meringkas arsip elektronik yang besar (Tong and Zhang 2016). "Topik" menandakan hubungan variabel yang tersembunyi, untuk diperkirakan, yang menghubungkan kata-kata dalam kosa kata dan kemunculannya dalam dokumen. Sebuah dokumen terdiri macam-macam topik campuran. Pemodelan topik ini dapat menemukan tema yang tersembunyi melalui koleksi dokumen dan dapat mengetahui keterangan dokumen sesuai dengan tema tersebut. Setiap kata dilihat dan sebagian diambil dari salah satu topik tersebut. Akhirnya, sebuah dokumen cakupan dari distribusi topik dihasilkan dan menyediakan cara baru untuk mengeksplorasi data tentang perspektif topik. Pemodelan topik dapat mengatur dan meringkas dokumen atau arsip elektronik pada skala yang besar yang tidak mungkin bisa dilakukan oleh manusia (Buenano-Fernandez et al. 2020). Secara konsep dan teori, alur algoritma pemodelan topik LDA dapat dilihat pada Gambar 2.10.



Gambar 2.10 Alur Algoritma LDA

Latent Dirichlet Allocation adalah Pemodelan topik probabilistik yang setiap dokumen direpresentasikan secara acak atau campuran lebih dari satu set topik laten dan masing-masing topik direpresentasikan sebagai distribusi atas kosa kata (Kusumaningrum et al. 2017).

LDA dapat dilihat sebagai dua proses, yaitu LDA sebagai proses generatif dan LDA sebagai proses inferensi. LDA sebagai proses generatif diimplementasikan saat kita mau buat corpus, seperangkat dokumen, di mana kita sudah tahu nilai distribusi kata untuk setiap topik (φ) dan proporsi topik untuk setiap dokumen (θ). Sebaliknya, LDA sebagai proses inferensi diterapkan ketika kami ingin mengidentifikasi variabel laten termasuk distribusi kata untuk setiap topik dan proporsi topik untuk setiap dokumen dan kami hanya memiliki satu set kata-kata sebagai variabel yang diamati Seperti yang digambarkan pada Gambar. 2.11.

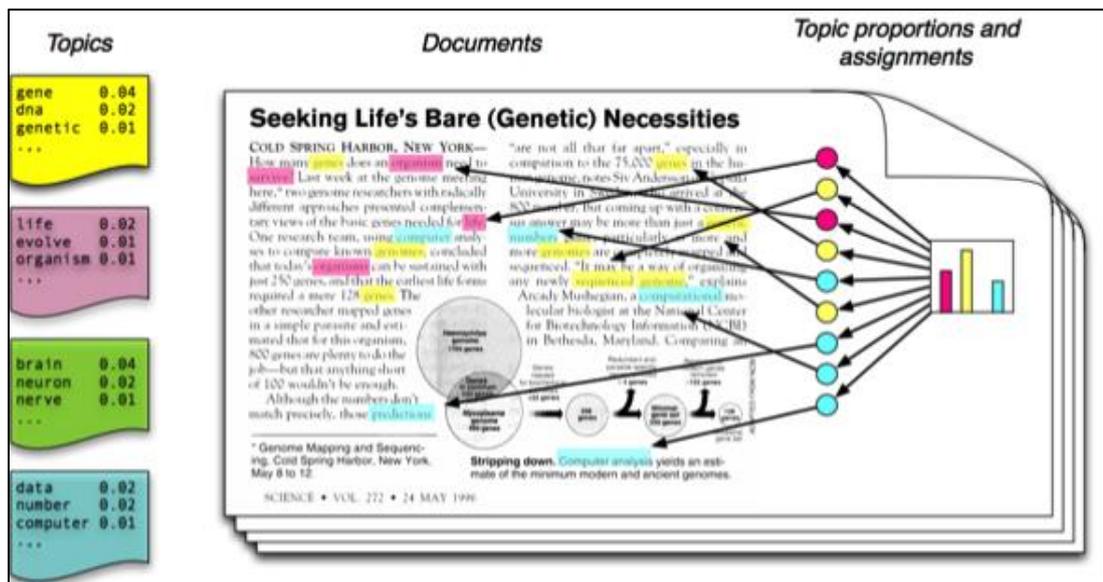


Gambar 2.11. Proses LDA

Pemodelan LDA direpresentasikan kedalam 3 tingkatan probabilitistik. Pertama lambda dan beta adalah parameter didalam tingkat *corpus*. Parameter lamda digunakan menentukan distribusi topik didalam dokumen. Dalam parameter tersebut bahwa semakin besar nilai lamda, maka semakin banyak juga topik yang dibahas. Adapun parameter beta digunakan untuk menentukan distribusi kata-kata yang berada dalam setiap topik. Dalam parameter tersebut menunjukan bahwa semakin kecil nilai beta maka semakin spesifik kata yang berada dalam topik. Untuk variabel θ adalah variabel dari tingkatan M dan dimana M ini menunjukan bahwa ada perulangan var sebanyak M kali untuk setiap dokumennya. Variabel θ ini digunakan untuk mempresentasikan distribusi topik pada dokumen tertentu saja. Hal ini menunjukan bahwa semakin kecil nilai θ , maka semakin spesifik juga topik yang terdapat dalam dokumen. Dalam variabel z_n dan w_n merupakan variabel dari tingkatan kata didalam dokumen N . Begitu pula dengan M, N menunjukan perulangan

pada variabel di dalamnya sebanyak N kali pada setiap kata. Untuk variabel z adalah variabel yang mempresentasikan topik dari kata tertentu pada dokumen. Untuk variabel w mempresentasikan topik dari kata tertentu pada dokumen, Sedangkan w variabel ini adalah variabel yang mempresentasikan yang berkaitan antara kata dengan topik tertentu yang berbeda pada sebuah dokumen.

Pemodelan dari *topic modeling* merupakan sebuah topik yang terdiri dari kata-kata tertentu yang dapat menyusun topik tersebut dari dokumen-dokumen yang memberikan suatu makna. Pemodelan topik pada penelitian ini yaitu digunakan untuk menemukan topik dalam publikasi penelitian kesehatan Indonesia berdasarkan judul jurnal. Pemodelan topik dapat menggambarkan sebuah makna dari dokumen secara semantik yang tersembunyi dalam teks yang jumlahnya besar dan dapat menemukan informasi dari data teks yang tidak terstruktur. Menurut Blei cara kerja LDA ditunjukkan pada Gambar 2.12.



Gambar 2.12 Cara kerja LDA

Pada Gambar 2.12 LDA menurut Blei mengasumsikan proses generatif berikut adalah rumus matematis untuk setiap dokumen w dalam sebuah *corpus* \mathbf{D} adalah sbb:

1. Pilih $N \sim \text{Poisson}(\xi)$,
2. Pilih $\theta \sim \text{Dir}(\alpha)$,
3. Untuk setiap N kata w_n ,
 - a. Pilih Topik $z_n \sim \text{Multinomial}(\theta)$,
 - b. Pilih sebuah kata w_n dari $p(w_n | z_n, \beta)$.

Keterangan:

- ξ : adalah *Poisson* kategorikal distribusi memakai variabel random diskrit
- Dir* : adalah *Dirichlet*

- M : adalah menunjukkan jumlah dokumen
 N : adalah jumlah kata dalam dokumen yang diberikan
 K : adalah jumlah topik
 W : adalah identitas semua kata dalam semua dokumen
 V : adalah jumlah kata dalam kosakata
 Z : adalah identitas topik dari semua kata dalam semua dokumen
 α : adalah parameter dari *Dirichlet* sebelum distribusi topik per-dokumen α_i
 β : adalah parameter dari *Dirichlet* sebelum distribusi kata per topik
 θ : adalah distribusi topik dokumen
 w^j : adalah identitas semua kata dalam semua dokumen j
 z^i : adalah topik untuk kata dalam dokumen i
 z^{ij} : adalah topik untuk kata ke-j dalam dokumen i
 w^{ij} : adalah kata spesifik
 θ_i : adalah distribusi topik untuk dokumen i
 θ_d : adalah level variabel dokumen dan sampel diambil sekali per-dokumen
 φ_k : adalah distribusi kata untuk topik k
 α_i : adalah parameter dari *Dirichlet* sebelum distribusi topik per-dokumen i
 α_k : adalah parameter dari *Dirichlet* jumlah topik per-dokumen k
 w_n : adalah identitas semua kata dalam semua dokumen n
 z_n : adalah identitas topik dari semua kata dalam semua dokumen n
 β_{ij} : adalah parameter dari *Dirichlet* ke-j dalam parameter per topik i
 w_{dn} : adalah *word-level* variabel dan sampel diambil sekali kata disetiap *doc*
 z_{dn} : adalah *topic-level* variabel dan sampel diambil sekali kata disetiap *doc*

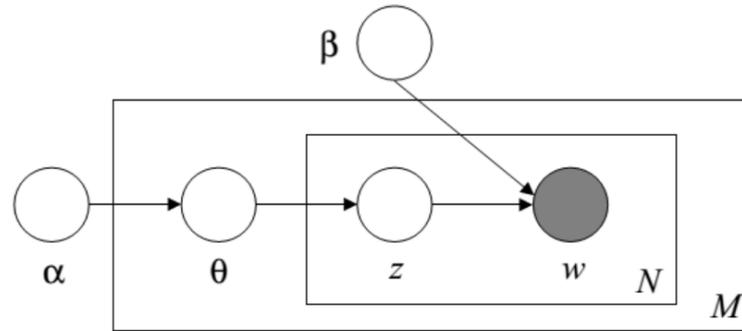
Beberapa asumsi penyederhanaan yang dibuat didalam distribusi dari (*latent*) topik bahwa diketahui mengikuti k distribusi *Dirichlet*. Kedua, probabilitas kata adalah matriks β berukuran $k \times V$ yang mana $\beta_{ij} = p(w^j = 1 | z^i = 1)$. Sedangkan k sebagai distribusi *Dirichlet* memiliki fungsi densitas dapat dilihat pada persamaan (1) sebagai berikut:

$$p(\theta | \alpha) = \frac{\Gamma(\sum_{i=1}^k \alpha_i)}{\prod_{i=1}^k \Gamma(\alpha_i)} \theta_1^{\alpha_1-1} \dots \theta_k^{\alpha_k-1} \quad (1)$$

Adapun bentuk dalam distribusi bersama dari Topik *mixture* θ dari N topik Z dan N kata w besyarat α dan β dapat dilihat pada persamaan (2) sebagai berikut:

$$p(\theta, \mathbf{z}, \mathbf{w} | \alpha, \beta) = p(\theta | \alpha) \prod_{n=1}^N p(z_n | \theta) p(w_n | z_n, \beta) \quad (2)$$

Bentuk dari representasi model LDA dapat digambarkan dalam sebuah diagram dapat dilihat pada Gambar 2.13.



Gambar 2.13 Representasi Model LDA

Bentuk distribusi marginal dari $p(\mathbf{w} | \alpha, \beta)$ didapat dengan menintegrasikan persamaan (2) terhadap θ dapat menghasilkan persamaan (3):

$$p(\mathbf{w} | \alpha, \beta) = \int (\theta | \alpha) \left(\prod_{n=1}^N \sum_{z_n} p(z_n | \theta) p(w_n | z_n, \beta) \right) d\theta \quad (3)$$

Akhirnya, diperoleh perkalian densitas marginal untuk sebuah dokumen yang akan memperoleh probabilitas marginal sebuah *corpus* persamaan (4) sebagai berikut:

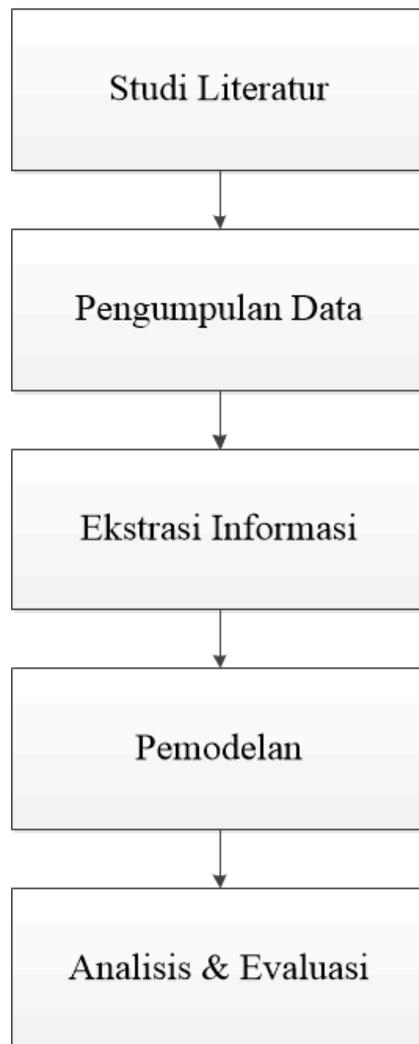
$$p(\mathbf{D} | \alpha, \beta) = \prod_{d=1}^M \int p(\theta_d | \alpha) \left(\prod_{n=1}^{N_d} \sum_{z_n} p(z_{dn} | \theta_{dn}, \beta) \right) d\theta_d \quad (4)$$

Visualisasi Pemodelan Topik pada penelitian ini untuk menghasilkan visualisasi pemodelan topik yang representatif dan mudah dipahami, penulis menggunakan *pyLDAvis* dan *word cloud*. *Wordcloud* merupakan sebuah sistem yang memunculkan susunan kata sebagai citra visual terkait frekuensi kemunculan kata dalam suatu teks verbal. Menurut (Kusumaningrum et al. 2017) berargumen bahwa visualisasi *wordcloud* dari teks akan memudahkan pengamat dalam melihat gagasan dan pendirian sang penulis teks sehingga dapat menjadi alat bantu dalam melakukan analisis terhadap sebuah wacana. *PyLDAvis* adalah dirancang untuk membantu pengguna menafsirkan topik dalam model topik yang sesuai dengan kumpulan data teks (Dunkel et al. 2019). *PyLDAvis* dapat digunakan untuk melihat sebaran kata untuk setiap topik beserta tingkat relevansinya yang dapat diranking dari nilai tertinggi hingga terkecil. *PyLDAvis* juga memvisualisasikan bagaimana kedekatan antara topik yang satu dengan yang lain menggunakan diagram kartesian. Sedangkan untuk *word cloud*, visualisasi tersebut akan menampilkan sebaran kata tiap topik.

BAB 3

Metodologi

Bab ini menjelaskan bagaimana cara penelitian dilakukan sehingga dapat diketahui rincian tentang urutan langkah-langkah yang dibuat secara sistematis dan dapat dijadikan pedoman yang jelas dalam menyelesaikan permasalahan, membuat analisis terhadap hasil penelitian, serta kesulitan-kesulitan yang dihadapi. Metodologi penelitian yang digunakan dalam penelitian ini sebagaimana ditunjukkan oleh Gambar 3.1



Gambar 3.1 Metode Penelitian

3.1 Studi Literatur

Pada Studi literatur dalam penelitian ini yaitu cara yang dipakai untuk menghimpun data-data atau sumber-sumber yang berhubungan dengan topik dalam suatu penelitian. Studi literatur didapatkan dari berbagai sumber, jurnal, buku dokumentasi, internet dan pustaka baik yang bersifat *online source* ataupun *offline source*.

Topik yang digali dalam langkah studi pustaka dalam penelitian ini adalah topik yang berkaitan dengan *text mining*, penelitian kesehatan, SINTA, *Network*, *web scarping*, *graph*, pemodelan topik, visualiasai serta informasi lain yang relevan dengan topik penelitian.

3.2 Pengumpulan Data

Pada tahap ini dilakukan pengumpulan data tentang topik penelitian. Semua tahap pada proses pengumpulan data tersebut diperoleh dari Website Jurnal SINTA yang berkaitan tentang Penelitian Kesehatan di Indonesia. Website jurnal tentang kesehatan tersebut dilakukan *scraping* untuk mengambil data dari jurnal bidang kesehatan berdasarkan keyword, judul, author, tahun, topik dan afiliasi. *Scraping* adalah adalah proses yang digunakan oleh mesin pencari (search engine) untuk mengumpulkan halaman website (Castillo, C., 2005) mengambil data dari jurnal tersebut yang dilakukan terfokus kepada penelitian kesehatan di Indonesia. Data yang dikumpulkan pada penelitian ini pada bulan Januari – Februari 2020.

3.3 Ekstrasi Informasi

Ekstraksi informasi dalam penelitian ini yaitu untuk pengolahan bahasa alami, dengan cara mengubah teks tidak terstruktur menjadi informasi dalam bentuk terstruktur (Susanti, E. and Mustofa, K., 2015). Berbagai jenis informasi di Internet ditransmisikan secara tidak terstruktur melalui website jurnal di Indonesia, menyebabkan munculnya kebutuhan akan suatu teknologi untuk menganalisa teks dan menemukan pengetahuan yang relevan dalam bentuk informasi terstruktur dengan menggunakan analisis *text mining*. *Text Mining* merupakan penambangan teks didefinisikan sebagai proses mengekstraksi pengetahuan implisit dari data tekstual (Onno W. Purbo, 2019). *Text mining* memberikan sebuah set metodologi dan tools untuk menemukan (*Discovering*), Memvisualisasikan (*presenting*), dan mengevaluasi pengetahuan dari kumpulan besar dari dokumen-dokumen teks (Data, Y. and Sarkar, D., n.d.). Sehingga *Tekt Mining* menurut terdapat 3 proses yang perlu dijalankan dalam *text mining* dapat dilihat pada Gambar 3.2.

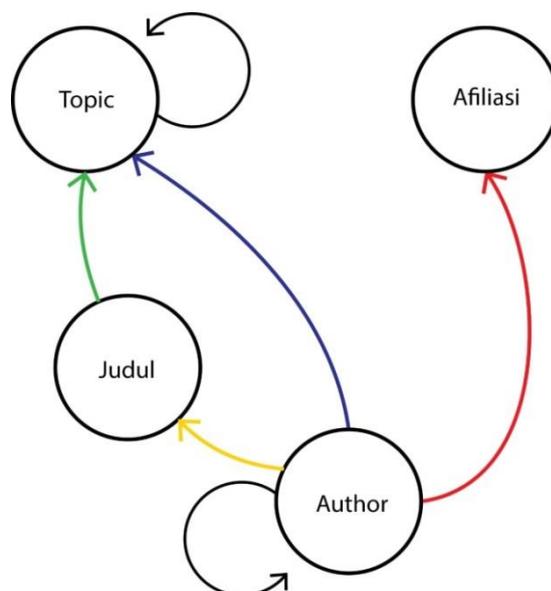


Gambar 3.2. Langkah Ekstrasi Informasi

Pada Gambar 3.2 langkah pertama dalam ekstrasi informasi pertama yaitu data *acquisition data*, pada tahap ini yaitu proses yang digunakan untuk mengumpulkan informasi untuk dokumen dan menganalisis beberapa fenomena. Pengumpulan data dalam penelitian ini diperoleh dari website Jurnal SINTA dengan cara *scraping*. Data yang digunakan dalam penelitian ini yaitu data yang berkaitan riset dibidang kesehatan di Indonesia. Langkah kedua yaitu *Text Processing* dimana data diperoleh yang semula tidak terstruktur kemudian dijadikan data yang terstruktur. Langkah ketiga yaitu membangun pemodelan untuk memodelkan dengan *network* atau jejaring penelitian yang digunakan untuk menganalisis mengetahui semua hubungan antara masing-masing pasangan publikasi dalam jaringan, tujuannya yaitu dapat diketahui pola yang dihasilkan dalam jejaring tersebut kemudian dapat divisualisasikan dalam bentuk *graph*.

3.4 Pemodelan

Pemodelan penelitian ini menggunakan Jejaring Model atau *Network analysis*.. Metode *Networking* dipilih karena metode ini dapat memberikan gambaran atau visualisasi sampai pada hubungan terkecil yang terjadi hanya pada satu individu dengan satu individu lainnya di dalam jaringan, metode ini juga dapat digunakan untuk menemukan *node*, *communities*, dan informal *hierarchies* yang memiliki pengaruh paling besar di dalam jaringan digunakan untuk menentukan *node* sentral didalam sebuah *network*. Pemodelan ini dibangun berdasarkan data yang diperoleh dari hasil *scraping* data pada SINTA dengan *keyword* kesehatan dan *health*. Data yang dihasilkan yaitu dari penelitan kesehatan yang terindeks SINTA berdasarkan topik, *author*, judul, tahun dan afiliasi. *Networking* yang diharapkan dalam penelitian yang akan dibangun dapat dilihat pada Gambar 3.3.



Gambar 3.3 *Networking* Analisis Publikasi

Pada Gambar 3.3 dibangun berdasarkan data yang diperoleh dari hasil *scraping* data pada SINTA dengan *keyword* kesehatan dan *health*. Data yang dihasilkan yaitu dari penelitian kesehatan yang terindeks SINTA berdasarkan topik, *author*, judul, tahun dan afiliasi. Pemodelan jejaring yang dibangun berisi topik penelitian kesehatan, judul-judul penelitian kesehatan, afiliasi penelitian kesehatan, *author*. Pemodelan pada Gambar 3.3 tersebut dalam penelitian ini digunakan untuk acuan pembuatan pemodelan jejaring penelitian kesehatan di Indonesia. Pemodelan jejaring yang dilakukan dalam penelitian ini diharapkan dapat mengetahui judul, kerjasama afiliasi, *author*, dan konsentrasi topik penelitian kesehatan di Indonesia sehingga dapat mengetahui secara mudah untuk dianalisis tren penelitian kesehatan di Indonesia.

3.5 Analisis dan Evaluasi

Dalam proses analisis data penelitian ini adanya diskusi menurut sumber dari kemenkes, WHO dan sumber lainnya dengan memperhatikan hasil jejaring penelitian berdasarkan topik, judul, *author*, afiliasi dan tahun yang dimodelkan. Hasil dari jejaring dan topik-topik penelitian kesehatan Indonesia yang sudah diimplementasikan kemudian dianalisis untuk memberikan sebuah makna terhadap data. Adapun analisis yang diharapkan dapat mengetahui penelitian-penelitian kesehatan per tahun, fokus penelitian kesehatan, fokus topik penelitian tertentu, dan afiliasi penelitian kesehatan di Indonesia. Hasil Analisis penelitian ini diharapkan dapat membantu Peneliti Kesehatan, Pelayan Kesehatan, Pemerintah dalam mengambil keputusan dalam hal memberikan pelayanan kesehatan di Indonesia. Penelitian ini membangun sebuah pengetahuan jejaring penelitian kesehatan di Indonesia yang kemudian dapat dianalisis lebih lanjut salah satu implementasinya dengan *topic modelling* yang digunakan untuk melihat tren suatu topik penelitian pada tahun tertentu. Metode *topic modeling* yang digunakan adalah LDA. Hasil dari *topic modelling* bertujuan untuk mengidentifikasi dan melakukan pemodelan topik terhadap judul-judul penelitian di bidang penelitian kesehatan Indonesia pada tahun-tahun tertentu. Pada tahap evaluasi pada penelitian ini yaitu dengan melihat penilaian tentang manfaat, keseimbangan dan nilai dari hasil penelitian dengan cara melihat hasil dari pemodelan pengetahuan dan pemodelan topik yang dihasilkan. Evaluasi dalam penelitian ini dengan mengevaluasi struktur sistem memenuhi seluruh komponen masukan, proses dan hasil. Fokus evaluasi pada penelitian ini adalah membangun pengetahuan jejaring penelitian kesehatan Indonesia berdasakan *author*, judul, topik, afiliasi, dan tahun dengan metode *text mining*, Memvisualiasaksikan dan menganalisis tren penelitian kesehatan indonesia berdasarkan judul penelitian kesehatan di Indonesia.

BAB 4

Hasil dan Pembahasan

Pada Bab ini menggambarkan mengenai implementasi perancangan yang sebelumnya telah dibahas pada Bab III. Pada struktural laporan di bab ini terdapat beberapa implementasi yang penulis lakukan yaitu pengumpulan data, hasil pemodelan pengetahuan jejaring penelitian kesehatan, hasil pemodelan topik penelitian kesehatan, dan analisis dan evaluasi.

4.1 Pengumpulan Data

Pengumpulan data menggunakan Implementasi teknik *web scraping* dengan menggunakan id jurnal penelitian kesehatan di Indonesia. Tahap menentukan target jurnal ini yaitu dengan mengakses Jurnal SINTA dan mencari menggunakan *keyword* atau kata kunci kesehatan dan *health* di Indonesia. Data diambil dengan menggunakan kata kunci yang khusus riset-riset bidang kesehatan. Teknik *web scraping* yang digunakan untuk menavigasikan dan memparser DOM (*Document Object Model*) untuk menentukan posisi judul, topik, afiliasi, dan *author*. Hasil pengumpulan data *web scraping* disimpan dalam bentuk *xlsx* yang kemudian dijadikan dalam bentuk format CSV. Adapun cuplikan hasil target Jurnal dapat dilihat pada Tabel 4.1. Data yang berhasil di *scraping* sebanyak data 17855 dari 172 jurnal penelitian kesehatan di Indonesia. Berikut hasil cuplikan hasil *scraping* disajikan pada Tabel 4.2.

Tabel 4.1. Cuplikan Hasil Target Nama Jurnal Penelitian Kesehatan di Indonesia

NO	JUDUL
1	Jurnal Ekologi Kesehatan
2	KEMAS : Jurnal Kesehatan Masyarakat
3	Media Penelitian dan Pengembangan Kesehatan
4	Jurnal Kesehatan Lingkungan Indonesia
5	Medisains : Jurnal Ilmiah Ilmu - Ilmu Kesehatan
6	Jurnal Administrasi Kesehatan Indonesia
7	Media Kesehatan Masyarakat Indonesia
8	Buletin Penelitian Kesehatan
9	Buletin Penelitian Sistem Kesehatan
10	Berkala Ilmu Kesehatan Kulit dan Kelamin (Periodical of Dermatology and Venerology)

Tabel 4.2 Contoh Hasil *Scraping*

No	Judul	Topik (<i>Keyword</i>)	Afiliasi	<i>Author1</i>	<i>Author2</i>	<i>Author3</i>
1	Pengetahuan dan persepsi masyarakat tentang malaria kaitannya dengan kondisi lingkungan di Kabupaten Banjarnegara	<i>knowledge, perception, malaria, endemic, environment</i>	Puslitbang Ekologi Kesehatan	S Suharjo	S Sukowati	H Manalu
2	Beberapa aspek bionomik vektor malaria di Sulawesi Tengah	<i>Bionomic, Vector, Malaria, Central Sulawesi</i>	Puslitbang Sulawesi	J Jastal	Y Wijaya	T Wibowo
3	Perilaku Masyarakat Dalam Menggunakan Kelambu Celup Di Daerah Endemik Malaria, Mimika Timur, Irian Jaya	<i>Anopheles barbirostris, malaria, mitotic chromosome, sibling species</i>	Puslitbang Ekologi Kesehatan	S Suharjo	SS Santoso	H Manalu

Pada Tabel 4.2 tersebut cuplikan hasil dari *scraping* yaitu dengan memperoleh judul, topik (*keyword* makalah), afiliasi, *author 1*, *author 2*, *author 3* dalam bentuk *xlsx* kemudian dijadikan kedalam bentuk format CSV untuk dimodelkan jejaring dengan *graph database*.

4.2 Hasil Pemodelan Jejaring

Implementasi pemodelan jejaring dalam penelitian ini menggunakan *graph database* Neo4j. *Query* yang digunakan dalam *database* Neo4j yaitu CQL (*Chyper Query Language*). CQL adalah *query* sebuah bahasa pemrograman yang digunakan oleh Neo4j untuk melakukan interaksi sebuah perintah yang digunakan dalam menjalankan sebuah program (Francis et al. 2018). Perintah CQL yang digunakan *create, match, return, where, marge, unique, load, delete*, dan lain-lain. Untuk mengimplementasikan pemodelan jejaring pengetahuan langkah pertama yaitu *load* data tesis dalam format csv. Setelah load data langkah selanjutnya yaitu membuat *node* judul, topik, afiliasi, *author1*, *author2*,

author3, dan tahun dengan *cypher query* dari data tesis yang sudah di *load*. Adapun cara untuk membuat masing- masing *node* ditunjukkan pada Gambar 4.1.

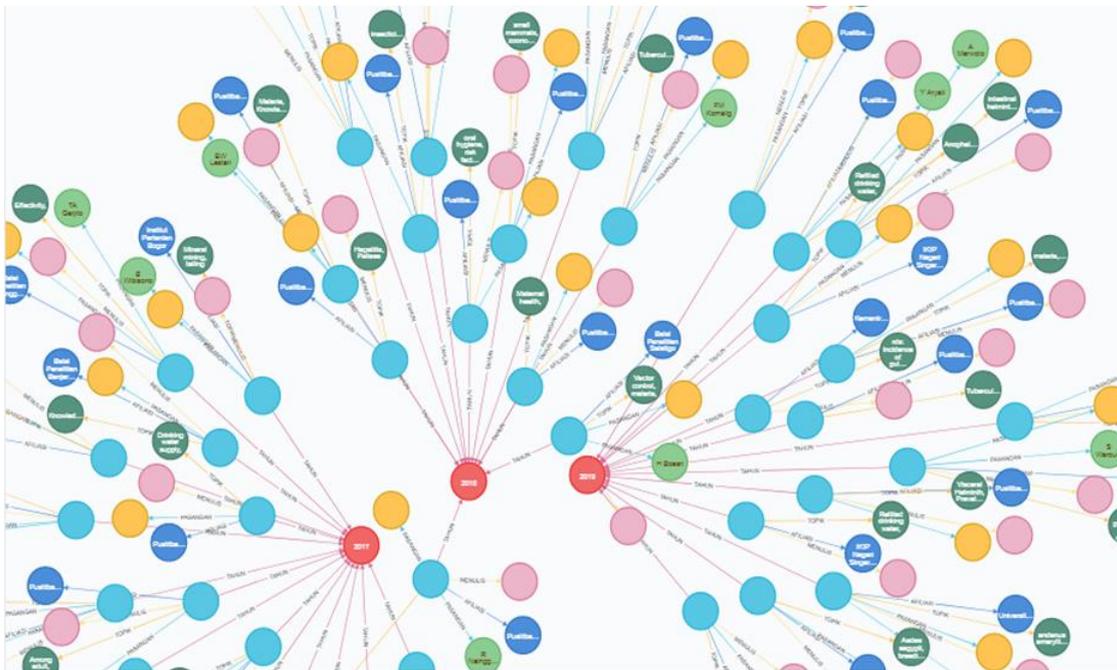
```
#Create Node
CREATE(A:Judul{judul: line.judul})
CREATE(B:Topik{Topik: line.Topik})
CREATE(C:Afiliasi{Afiliasi: line.Afiliasi})
CREATE(D:Author1{Author1: line.Author1})
CREATE(E:Author2{Author2: line.Author2})
CREATE(F:Author3{Author3: line.Author3})
CREATE(G:Tahun{Tahun: line.Tahun})
```

Gambar 4.1 Membuat *Node*

Setelah terbentuk semua *node* langkah selanjutnya merelasikan jejaring penelitian dengan masing-masing *node* yang terbentuk. Berikut cuplikan *cypher query* ditunjukkan pada Gambar 4.2 dan hasil pemodelan pengetahuan jejaring penelitian kesehatan Indonesia dapat dilihat pada Gambar 4.3.

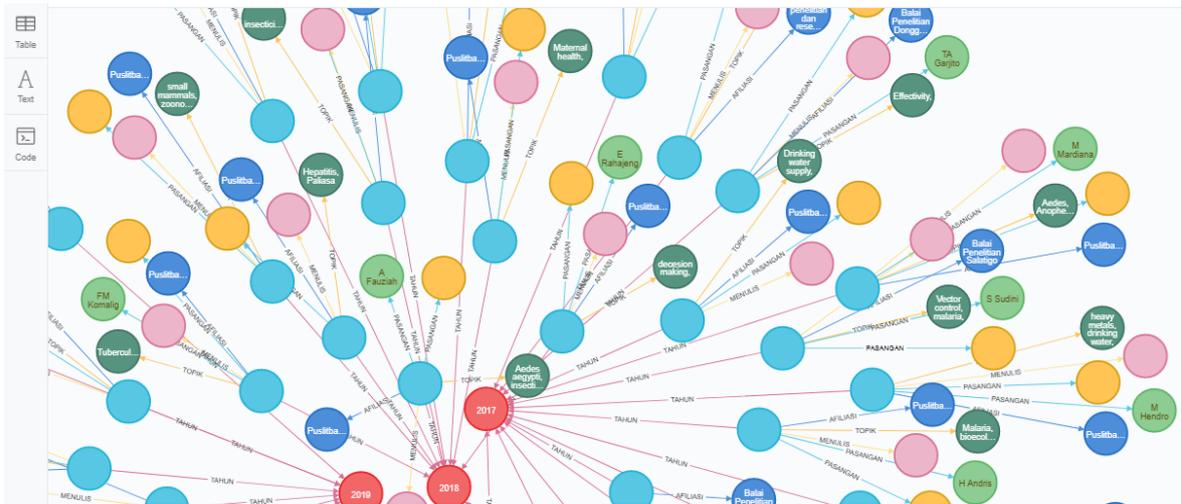
```
#Create Relation From Node
create (D)-[:PASANGAN{Author2:"Author2"}]->(E)
create (D)-[:PASANGAN{Author3:"Author3"}]->(F)
create (D)-[:MENULIS{judul:"Judul"}]->(A)
create (D)-[:TOPIK{Topik:"Topik"}]->(B)
create (D)-[:AFILIASI{Afiliasi:"Afiliasi"}]->(C)
create (D)-[:TAHUN{Tahun:"Tahun"}]->(G)
#Show Relation ALL
MATCH n Return n
```

Gambar 4.2 *Query* antar *Node*



Gambar 4.3 Hasil Pemodelan Jejaring Penelitian Kesehatan di Indonesia

Gambar pemodelan pengetahuan jejaring penelitian kesehatan yang lebih jelas dapat dilihat pada Gambar 4.4.



Gambar 4.4. Perbesar Hasil Pemodelan

Pada Gambar 4.4 tersebut hasil dari pemodelan jejaring penelitian kesehatan di Indonesia yang divisualisaikan dengan *node*, properti dan label. Untuk *node* yang berwarna merah merupakan *node* untuk tahun, warna ungu untuk judul, warna hijau tua untuk topik, warna biru tua untuk Afiliasi, warna biru muda untuk *Author* 1, Warna Orange Untuk *Author* 2 dan Warna hijau untuk *Author* 3. Untuk properti yang ditampilkan pada Gambar 4.4 yaitu sebuah isi yang terdapat di dalam masing-masing *node* sebagai contoh *node* berwarna ungu didalam *node* tersebut berisi judul-judul penelitian kesehatan Indonesia. Untuk label pada hasil pengetahuan ini menggunakan 5 label yaitu menulis, bekerjasama, topik, berafiliasi, dan tahun. Hasil Implementasi pemodelan jaringan Gambar 4.4 berhasil dimodelkan dimana dapat mengetahui jejaring penelitian kesehatan di Indonesia. Pemodelannya dapat mengetahui *Author* menulis judul tentang apa, *Author* berkerja sama dengan *Author* yang lainnya, kemudian dapat mengetahui *Author* berafiliasi dari mana, dapat melihat topik apa saja yang diteliti, dan pada tahun berapa *author* meneliti penelitian kesehatan di Indonesia.

4.3 Hasil Analisis Pemodelan Pengetahuan

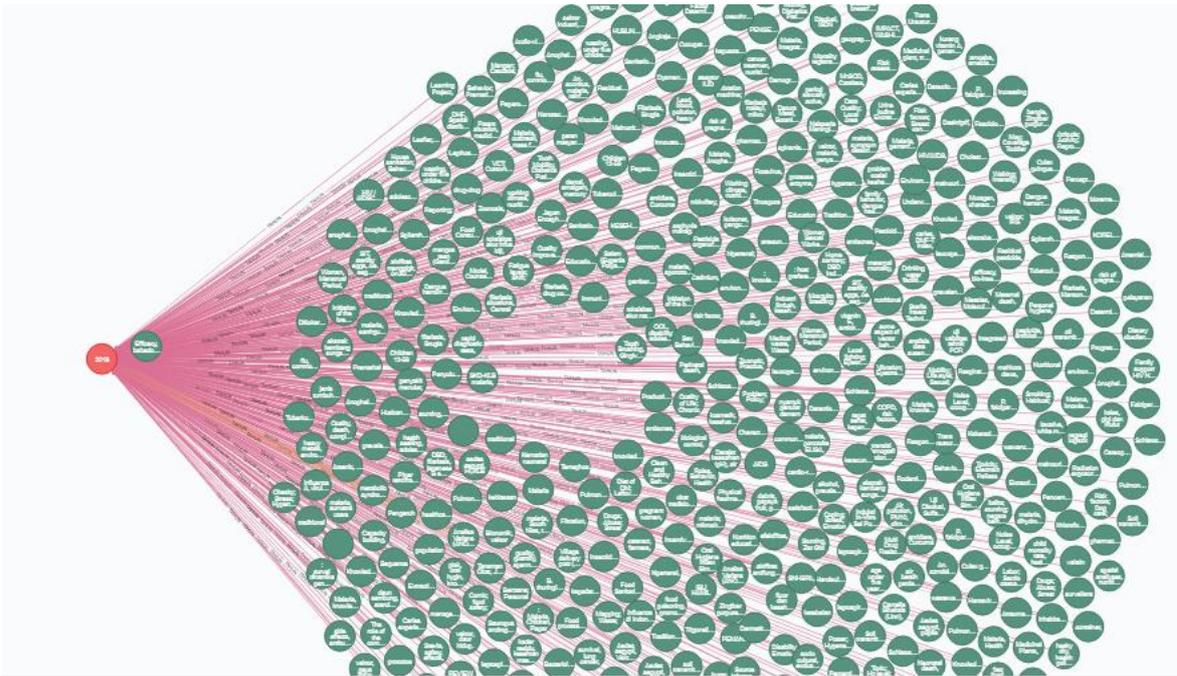
Dari hasil pemodelan pengetahuan yang dihasilkan pada Gambar 4.3 kita bisa analisis melihat kondisi penelitian medis di Indonesia dengan *query*. *Query* yang dapat diterapkan dapat melihat posisi riset di Indonesia. Adapun *query* yang bisa diambil dari hasil pemodelan pengetahuan dapat melihat topik-topik berdasarkan tahun, melihat judul riset berdasarkan tahun tertentu, melihat topik-topik berdasarkan afiliasi, melihat

judul berdasarkan topik, melihat *author* berdasarkan topik, melihat daerah-daerah berdasarkan topik penelitian di Indonesia. Dalam penelitian ini informasi atau *knowledge* yang didapatkan dan diturunkan yang bermanfaat dalam domain medis yaitu dapat mengetahui topik-topik apa saja yang pernah diteliti khususnya penelitian kesehatan Indonesia berdasarkan tahun adapun cuplikan *query* sebagai berikut *USING DATA TESIS LOAD CSV WITH HEADER FROM "dataset.csv" As Row Create(:Topik{Topik:row.Topik, :Tahun(Tahun:row.Tahun)}); CREATE ON; Topik(Topik); CREATE ON; Tahun(Tahun); create(Topik)-[:TAHUN{Tahun:"Tahun"}]->(Tahun) Match n Return n limit 3000* dan hasilnya ditunjukkan pada Gambar 4.5.



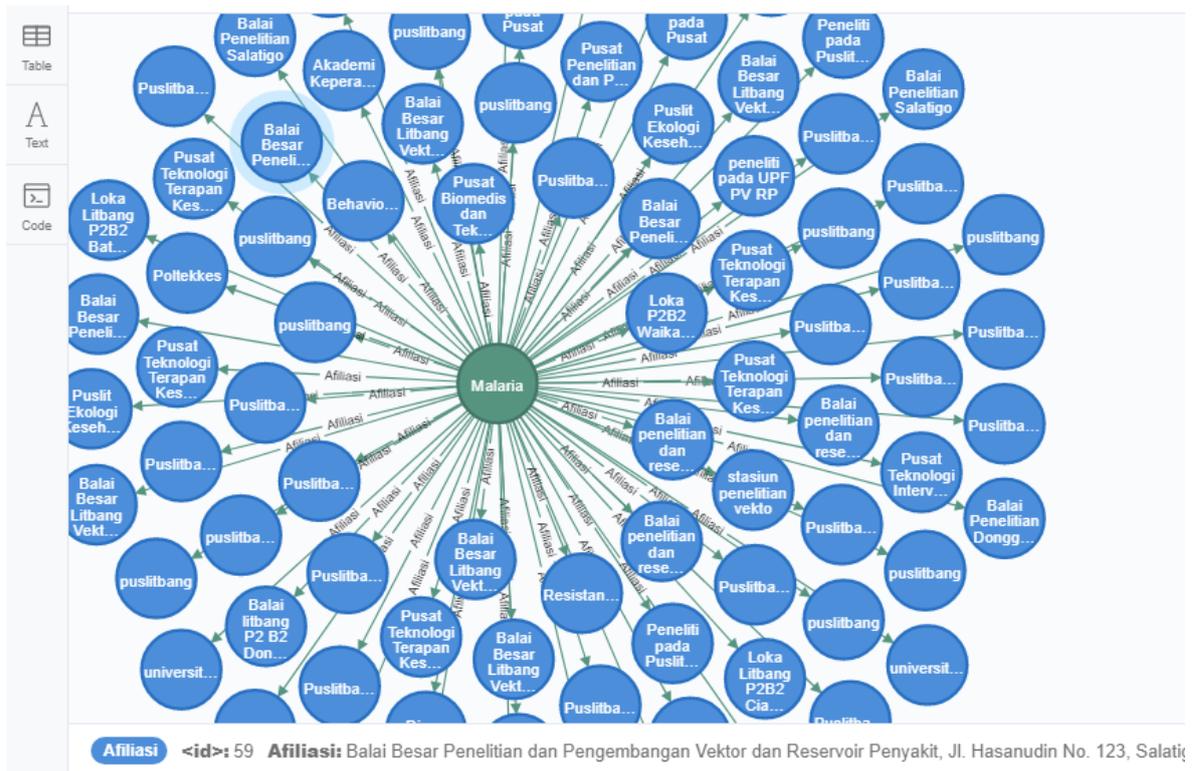
Gambar 4.5 Topik – Topik Berdasarkan Tahun

Dari Gambar 4.5 topik-topik penelitian kesehatan di Indonesia berdasarkan tahun untuk dapat diperjelas dapat dilihat pada Gambar 4.6.



Gambar 4.8 Topik-Topik Penelitian Kesehatan Pada Tahun 2017

Dengan adanya *graph* relasi topik pada gambar 4.3 informasi atau *knowledge* yang didapatkan dan diturunkan kembali yang bermanfaat dalam domain medis yaitu dapat melihat topik yang paling banyak diteliti. Adapun dari implementasi *graph* topik pada Gambar 4.3 didapatkan bahwa malaria adalah topik yang paling banyak diteliti. Setelah itu peneliti dapat melihat topik malaria berafiliasi dari mana saja yang pernah meneliti topik malaria. Berikut cuplikan *cypher query* untuk mengetahui afiliasi yang topiknya penyakit malaria sebagai berikut `LOAD CSV WITH HEADERS FROM "dataset.csv" AS row MATCH (B:Topik {Topik:row.Topik}) MATCH (C:Afiliasi{Afiliasi:row.Afiliasi}) UNIQUE (B: Topik {Topik:row. "Malaria:} MERGE (B)-[r:Afiliasi]->(c) RETURN *`; Adapun hasil relasi topik dengan afiliasi tersebut dapat ditunjukkan Gambar 4.9.



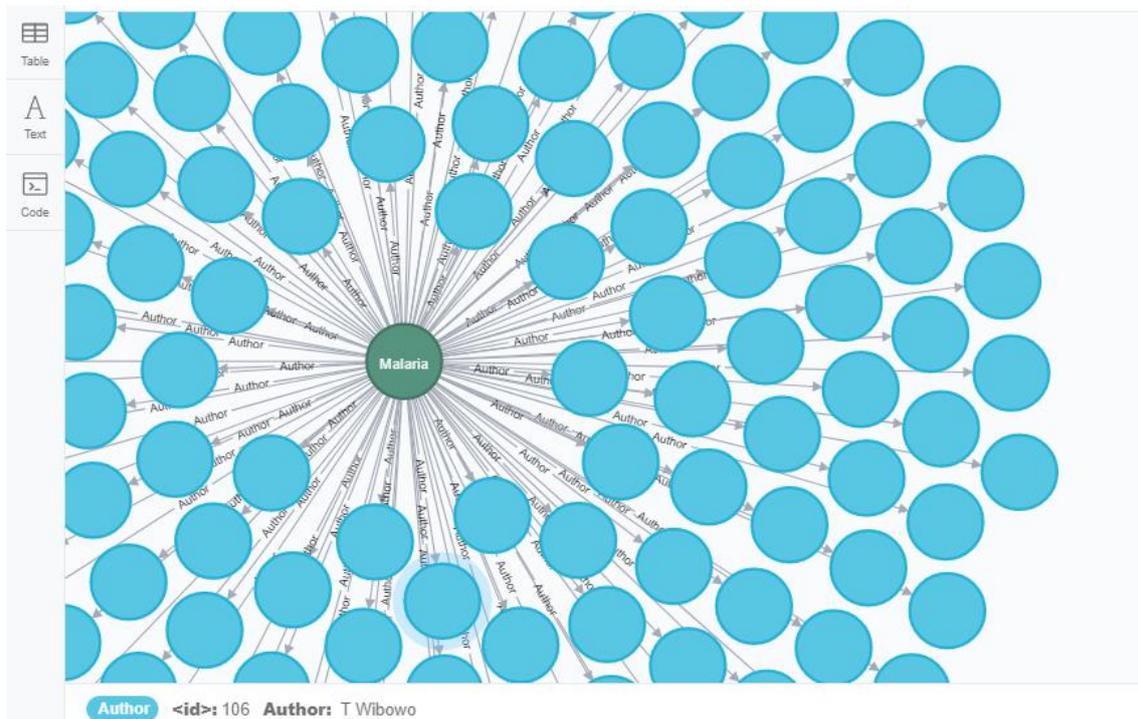
Gambar 4.9 Topik Riset Malaria dan Afiliasinya

Pada Gambar 4.9 tersebut kita dapat mengetahui topik penyakit malaria tersebut dari mana saja afiliasinya. Hal ini terlihat pada Gambar 4.9 contohnya afiliasi Balai Besar Penelitian dan Pengembangan Vektor dan Reservoir Penyakit, sedangkan untuk afiliasi yang lainnya dapat diketahui dengan mengklik masing-masing *node* warna biru tersebut. *Node* yang diklik bertujuan untuk melihat afiliasinya nya dari mana saja, sehingga keluar nama afiliasinya dari topik malaria.

Selain dapat mengetahui relasi topik malaria dan afiliasinya, informasi *graph* yang diimplementasikan dapat diturunkan kembali yaitu tujuannya untuk mengetahui *author* siapa saja yang meneliti topik penyakit malaria. Sehingga adanya relasi topik malaria dengan *author* dapat diketahui siapa saja yang aktif dalam meneliti topik malaria. Adapun cuplikan *cypher query* sebagai berikut *LOAD CSV WITH HEADERS FROM*

```
"Datatesis.csv" AS row MATCH (B:Topik {Topik:row.Topik}) MATCH (D:Author1 {Author1:row.Author1}) MATCH (E:Author2 {Author2:row.Author2}) MATCH (F:Author3 {Author3:row.Author3}) UNIQUE (B: Topik {Topik:row. "Malaria:}) MERGE (D)-[r:Author2]->(E)-[r:Author3]->(F) CREATE (B)-[r:Author]->(D) MATCH n RETURN n*;
```

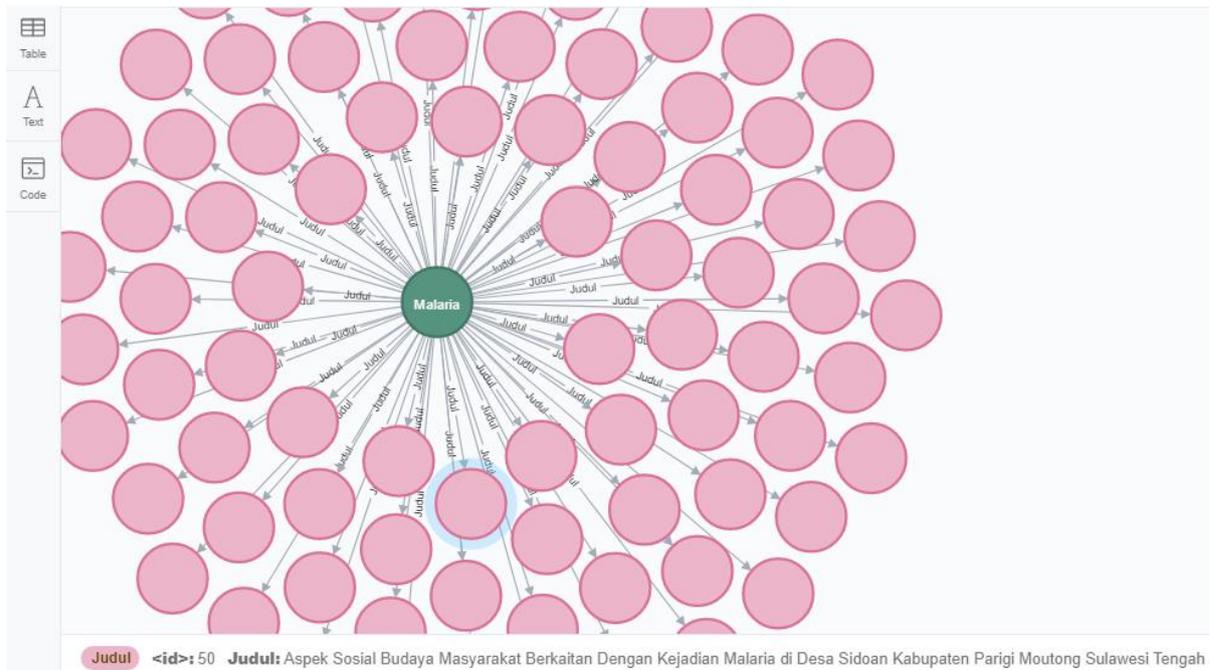
Adapun hasil dari *query* relasi topik malaria dan *author* ditunjukkan pada Gambar 4.10.



Gambar 4.10 Penelitian Topik Malaria dan *Author*

Pada Gambar 4.10 tersebut kita dapat mengetahui topik penyakit malaria dengan mengetahui relasi dari nama *author* siapa saja yang pernah meneliti topik malaria, hal ini terlihat pada Gambar 4.10 salah satu contohnya yang pernah meneliti topik malaria yaitu *author* bernama M Mulyono. Untuk melihat *author* yang lainnya dapat diketahui dengan mengeklik masing-masing *node* tersebut. Sedangkan untuk melihat *author* lainnya siapa saja yang pernah meneliti dengan mengeklik *node* warna biru, sehingga akan keluar nama *author*nya.

Dengan menggunakan metode *cypher query language* yang sama dapat diturunkan juga untuk mendapatkan informasi yang bermanfaat dalam domain medis yaitu judul-judul penelitian apa saja yang berkaitan dengan topik malaria. Adapun hasil *query* yang menunjukkan judul-judul penelitian yang berkaitan dengan topik malaria dapat dilihat pada Gambar 4.11.



Gambar 4.11 Judul Penelitian yang berkaitan dengan Malaria

Pada Gambar 4.11 hasil dari *query* dapat mengetahui judul-judul yang berkaitan dengan penyakit malaria. Dalam gambar tersebut judul apa saja yang berkaitan dengan topik malaria ini terlihat pada Gambar 4.11 salah satu contoh judul yang berkaitan dengan malaria “Persepsi dan Pola Kebiasaan Masyarakat Kaitannya dengan Masalah Malaria di Daerah Sihpeng Kabupaten Tapanuli Selatan Provinsi Sumatra Utara”. Untuk melihat judul-judul yang lainnya yang dapat diketahui dengan megeklik masing-masing *node*. Sedangkan untuk melihat judul-judul yang berkaitan dengan malaria lainnya caranya dengan mengeklik *node* warna merah muda, sehingga akan keluar judulnya.

4.4 Implementasi Pemodelan Topik

Implementasi pemodelan topik penelitian ini tujuannya adalah untuk mengetahui tren penelitian kesehatan di Indonesia. Setelah membangun pengetahuan jejaring penelitian kesehatan di Indonesia kemudian dapat dianalisis lebih lanjut salah satu implementasinya dengan *topic modelling*. Pemodelan topik yang dibangun dapat digunakan untuk melihat tren suatu topik penelitian pada tahun tertentu dengan memanfaatkan *query-query* yang dihasilkan. *Query* yang dihasilkan seperti judul berdasarkan tahun, judul berdasarkan afliasainya dan lain-lain, yang kemudian dapat dianalisis lebih lanjut dengan *topic modeling*. Implementasi untuk melihat tren topik penelitian kesehatan di Indonesia ini menggunakan salah satu metode *text mining* yaitu *topic modelling* LDA (*Latent Directed Allocation*). Dalam melakukan tahap mengimplementasikan pemodelan topik peneliti menggunakan bahasa pemrograman Python 3. Hasil dari *topic modelling* dilakukan dengan

tujuan untuk mengidentifikasi tren topik terhadap judul-judul penelitian di bidang penelitian kesehatan Indonesia. Adapun langkah-langkah pemodelan topik sebagai berikut:

4.4.1 *Preprocessing*

Preprocessing merupakan langkah yang cukup penting yang dapat berpengaruh pada hasil pemodelan. Penerapan langkah-langkah yang sesuai di dalam *preprocessing* diharapkan dapat meningkatkan kualitas pemodelan topik yang dihasilkan. Tujuan pada tahap *preprocessing* dalam penelitian ini adalah untuk membersihkan simbol, tanda baca, imbuhan, dan kata-kata yang tidak memiliki makna yang penting dan tidak memiliki pengaruh yang signifikan dalam melakukan interpretasi dari judul penelitian kesehatan. Pertama yaitu tahap *preprocessing* dengan beberapa proses yang harus dijalankan. Adapun pada penelitian ini dilakukan 5 tahap secara urut, di antaranya yaitu:

a. Menghapus tanda baca

Pada tahap menghapus tanda baca dilakukan untuk menghapus tanda baca atau simbol-simbol tertentu, seperti ., - ()&/'":+?!*#\$\$% dan lain sebagainya. Berikut cuplikan kode program untuk tahap menghapus tanda baca dapat ditunjukkan pada Gambar 4.12.

```
str = re.sub(r'[^a-z ]*([.0-9])*\d', '', str)
```

Gambar 4.12 Cuplikan kode tahap menghapus data

b. Menghapus angka

Pada tahap ini dilakukan untuk menghapus karakter berbentuk angka di antara spasi, namun tidak dengan angka yang berada di tengah atau berada langsung setelah atau sebelum alfabet. Sebagai contoh, tahap ini akan menghapus teks seperti “123” tetapi tidak menghapus teks “CH2”. Hal ini dikarenakan ada beberapa data judul penelitian yang mengandung nama metode atau obyek yang terdiri dari kombinasi huruf dan angka seperti “CH2”, “H20”, dan lain sebagainya. Berikut cuplikan kode program untuk tahap menghapus angka dapat ditunjukkan pada Gambar 4.13.

```
str = re.sub("^\d+\s|\s\d+\s|\s\d+$", " ", str)
```

Gambar 4.13 Cuplikan kode tahap menghapus angka

c. *Case Folding*

Pada tahap *case folding* dalam penelitian ini adalah untuk menyeragamkan bentuk huruf dari keseluruhan data teks menjadi huruf kecil. Berikut cuplikan kode program untuk tahap menghapus case folding ditunjukkan pada Gambar 4.14.

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

Gambar 4.14. Cuplikan kode *Case Folding*

d. Menghapus kalimat

Pada tahap ini dilakukan untuk menghapus baris data yang terdiri kurang dari atau sama dengan tiga kata. Umumnya, pada judul penelitian setidaknya terdiri lebih dari atau sama dengan 4 kata. Saat data sudah berhasil dikumpulkan menggunakan aplikasi, ditemukan beberapa baris yang tidak jelas dan tidak dapat dipahami sebagai judul penelitian. Contoh baris data yang ditemui di antaranya adalah “Daftar Isi”, “Indeks Nama Penulis”, dan lain sebagainya. Implementasi kode program untuk tahap ini dapat dilihat pada Gambar 4.15.

```
def removeSentence(str):  
    word = str.split()  
    wordCount = len(word)  
    if(wordCount<=3):  
        str = ''  
    return str
```

Gambar 4.15. Cuplikan kode menghapus kalimat kurang dari 3 kata

e. Menghapus *Stopwords*

Pada tahapan ini dilakukan untuk menghapus *stopwords*, Pada tahap ini yaitu dengan adanya kata-kata yang tidak begitu penting dan dianggap tidak memiliki pengaruh yang signifikan terhadap proses pemodelan topik pada penelitian yang akan dilakukan. Tahap ini dilakukan dengan mengumpulkan kata-kata tersebut kemudian disimpan dalam file dengan format *txt*, kemudian disimpan ke dalam folder *corpora* di *nlTK_data*. Adapun untuk setiap baris data dilakukan proses *tokenisasi* terlebih dahulu menjadi potongan kata tunggal untuk kemudian dilakukan pengecekan. Apabila terdapat kata yang teridentifikasi merupakan *stopwords*, maka kata tersebut akan dihapus. Berikut cuplikan kode program untuk tahap proses *stopwords* ditunjukkan pada Gambar 4.16.

```
def removeSentence(str):  
    word = str.split()  
    wordCount = len(word)  
    if(wordCount<=3):  
        str = ''  
    return str
```

Gambar 4.16. Cuplikan kode *stopwords*

4.4.2 Identifikasi Frasa

Tahap identifikasi frasa ini diawali dengan mempersiapkan data teks hasil *preprocessing* dari file dengan format *xls* ke dalam bentuk senarai. Implementasi kode program untuk proses ini dapat dilihat pada Gambar 4.17.

```
text = df['text']  
text_list = [i.split() for i in text]
```

Gambar 4.17. Tahap Identifikasi Frasa

Proses identifikasi frasa bertujuan untuk menggabungkan semua kata dalam data ke dalam bentuk *bigram* (dua kata) dan *trigram* (tiga kata) kemudian mencari kata yang digabungkan tersebut dengan kriteria kemunculannya. Setiap *bigram* dan *trigram* tersebut akan dianggap sebagai sebuah frasa. Frasa tersebut kemudian dimasukkan ke dalam satu kalimat. Implementasi kode program untuk membuat model *bigram* dan *trigram* dapat dilihat pada Gambar 4.18.

```
#Create Biagram & Trigram Models
from gensim.models import Phrases
# Add bigrams and trigrams to docs, minimum count 5 means only that appear
5 times or more.
bigram = Phrases(text_list, min_count=5)
trigram = Phrases(bigram[text_list], min_count=5)

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 4.18. Kode model *bigram* dan *trigram*

4.4.3 Perhitungan Bobot Setiap Kata

Pada tahap ini, terlebih dahulu dilakukan pembuatan dictionary yang merepresentasikan isi dari dokumen yang telah dilakukan *preprocessing*. Dalam membuat *dictionary*, penulis menggunakan modul *corpora.Dictionary()* dari *library Gensim* dengan parameter berupa data berbentuk dua dimensi yang berisi *array* dari *array* kata untuk setiap judul penelitian kesehatan di Indonesia. Implementasi kode program untuk membuat *dictionary* dapat dilihat pada Gambar 4.19.

```
# create dictionary
from gensim import corpora, models
dictionary = corpora.Dictionary(text_list)
```

Gambar 4.19. Kode program untuk membuat dictionary

Setelah *dictionary* berhasil dibuat, tahap selanjutnya adalah membuat *bag of words* dengan menerapkan fungsi *doc2bow* yang mengonversi dokumen menjadi format matriks dengan menghitung jumlah kemunculan kata yang unik. Implementasi kode program untuk membuat *bag of words* ditunjukkan pada gambar 4.20.

```
#build bag of words (corpus)
bow = [dictionary.doc2bow(doc) for doc in text_list]
```

Gambar 4.20. Cuplikan Kode Program *bag of words*

Setelah membuat *bag of words*, barulah dapat dilakukan penghitungan bobot untuk setiap kata. Penghitungan bobot dilakukan dengan menerapkan konsep TF-IDF. Dalam *Python*, untuk menghitung nilai tersebut dapat dilakukan dengan menggunakan dengan *library Gensim*. Implementasi kode program untuk menghitung nilai TF-IDF dapat dilihat pada gambar 4.21 sebagai berikut:

```
# calculate TF-IDF
tfidf = models.TfidfModel(bow)
corpus_tfidf = tfidf[bow]
```

Gambar 4.21 Kode program untuk menghitung nilai TF-IDF

4.4.4 Pemodelan Topik

Data yang digunakan pada pemodelan topik ini adalah judul-judul penelitian kesehatan Indonesia yang dihasilkan dari *query*. Data yang digunakan yaitu judul-judul penelitian pada tahun 2017, 2018 dan 2019. Pemodelan topik ini bertujuan untuk melihat topik-topik apa saja yang telah diteliti di Indonesia. Judul-judul yang dimodelkan dalam pemodelan topik ini terdapat 2 kategori yaitu judul secara keseluruhan dan judul pada tahun-tahun tertentu. Pada langkah pemodelan topik ini yaitu bertujuan untuk membuat model LDA untuk mengetahui tren topik penelitian kesehatan di Indonesia. Dalam membuat model LDA, penulis memanfaatkan *library Gensim* dengan fungsi *LdaModel()*, sedangkan untuk mendapatkan nilai koherensi, penulis menggunakan fungsi *CoherenceModel()*. Implementasi *fungsi LdaModel()*, terdapat 4 parameter yang digunakan, yakni *corpus*, *id2word*, *num_topics*, dan *iterations*. Sedangkan fungsi *CoherenceModel()*, terdapat 6 parameter yang digunakan, yakni *dictionary*, *corpus*, *texts*, *limit*, *start*, dan *step*. Adapun implementasi kode program dalam melakukan pembentukan model LDA penghitungan nilai koherensi dapat dilihat pada Gambar 4.22.

```

import numpy as np
from gensim.models.coherencemodel import CoherenceModel
from gensim.models.ldamodel import LdaModel
from gensim.corpora.dictionary import Dictionary
from numpy import array
#function to compute coherence values
def compute_coherence_values(dictionary, corpus, texts, limit, start,
step):
    coherence_values = []
    np.random.seed(9)
    model_list = []
    num_topics = []
    for num_topics in range(start, limit, step):
        model = LdaModel(corpus=corpus, id2word=dictionary,
num_topics=num_topics, iterations=100)
        model_list.append(model)
        coherencemodel = CoherenceModel(model=model, texts=texts,
dictionary=dictionary, coherence='c_v')
        coherence_values.append(coherencemodel.get_coherence())

    return model_list, coherence_values

start=2
limit=20
step=2
model_list, coherence_values = compute_coherence_values(dictionary,
corpus=corpus_tfidf,
texts=text_list,
start=start, limit=limit, step=step)

```

Gambar 4.22. Cuplikan Kode Pemodelan Topik LDA

Adapun rincian nilai koherensi untuk setiap banyak topik dapat dilihat dengan mengimplementasikan dengan hasil keluaran yang dapat dilihat pada Gambar 4.23.

```

[(0,
'0.048*"knowledge" + 0.037*"obat" + 0.037*"medicine" + 0.033*"anak" + '
'0.031*"jakarta" + 0.029*"tanaman" + 0.026*"nutritional_status" + '
'0.025*"risk" + 0.024*"children" + 0.023*"hygiene"'),
(1,
'0.058*"malaria" + 0.033*"traditional" + 0.032*"nutritional_status" + '
'0.027*"knowledge" + 0.025*"jawa" + 0.025*"treatment" + 0.022*"extract" + '
'0.021*"index" + 0.020*"water" + 0.018*"blood"'),
(2,
'0.060*"filariasis" + 0.053*"of" + 0.028*"kokap" + 0.026*"mosquito" + '
'0.026*"health" + 0.025*"dhf" + 0.022*"community" + 0.021*"local" + '
'0.021*"thuringiensis" + 0.021*"dbd"'),
(3,
'0.077*"aedes_aegypti" + 0.065*"kesehatan" + 0.033*"behavior" + '
'0.030*"aedes" + 0.025*"environment" + 0.025*"aegypti" + 0.024*"quality" + '
'0.023*"dengue" + 0.021*"health" + 0.019*"dbd"'),
(4,
'0.063*"disease" + 0.043*"anopheles" + 0.039*"pestisida" + 0.037*"activity" + '
'+ 0.036*"food" + 0.034*"uji" + 0.028*"vector" + 0.025*"pesticide" + '
'0.021*"blood" + 0.021*"of"'),
(5,
'0.113*"malaria" + 0.033*"health" + 0.031*"of" + 0.027*"anopheles" + '
'0.025*"management" + 0.024*"air" + 0.023*"vektor" + 0.021*"age" + '
'0.019*"pelayanan" + 0.019*"fungi"')]]

```

Gambar 4.23. Nilai Koherensi

Dari model pada Gambar 4.23 tersebut dapat melakukan inferensi topik disetiap judul yang ada di *corpus* menggunakan pemodelan topik yaitu Tf-Idf (*Term Frequency dan Invers*

Document Frequency). Inferensi ini dilakukan untuk mendapatkan distribusi topik dalam setiap judul. Setelah dilakukan pemodelan pada gambar tersebut selanjutnya mencari *coherence num topic* yang hasil dari *coherence value* atau nialainya yang paling tinggi. Berikut *num topic coherence value* yang paling tinggi dapat dilihat pada Gambar 24.

```
Topic: 0 Word: 0.042*"health" + 0.035*"of" +
0.025*"filariasis" + 0.023*"water" + 0.022*"disease" +
0.022*"nutritional_status" + 0.022*"obat" +
0.020*"breastfeeding" + 0.018*"care" + 0.017*"falciparum"
Topic: 1 Word: 0.035*"kesehatan" + 0.023*"air" +
0.021*"traditional" + 0.021*"anopheles" +
0.020*"management" + 0.018*"blood" + 0.018*"risk" +
0.018*"pregnancy" + 0.018*"perilaku" + 0.017*"malaria"
Topic: 2 Word: 0.080*"malaria" + 0.050*"aedes_aegypti" +
0.037*"knowledge" + 0.023*"aegypti" + 0.021*"vector" +
0.020*"aedes" + 0.018*"dhf" + 0.017*"food" +
0.017*"anopheles" + 0.017*"of"
```

Gambar 4.24. Nilai Topik Koherensi

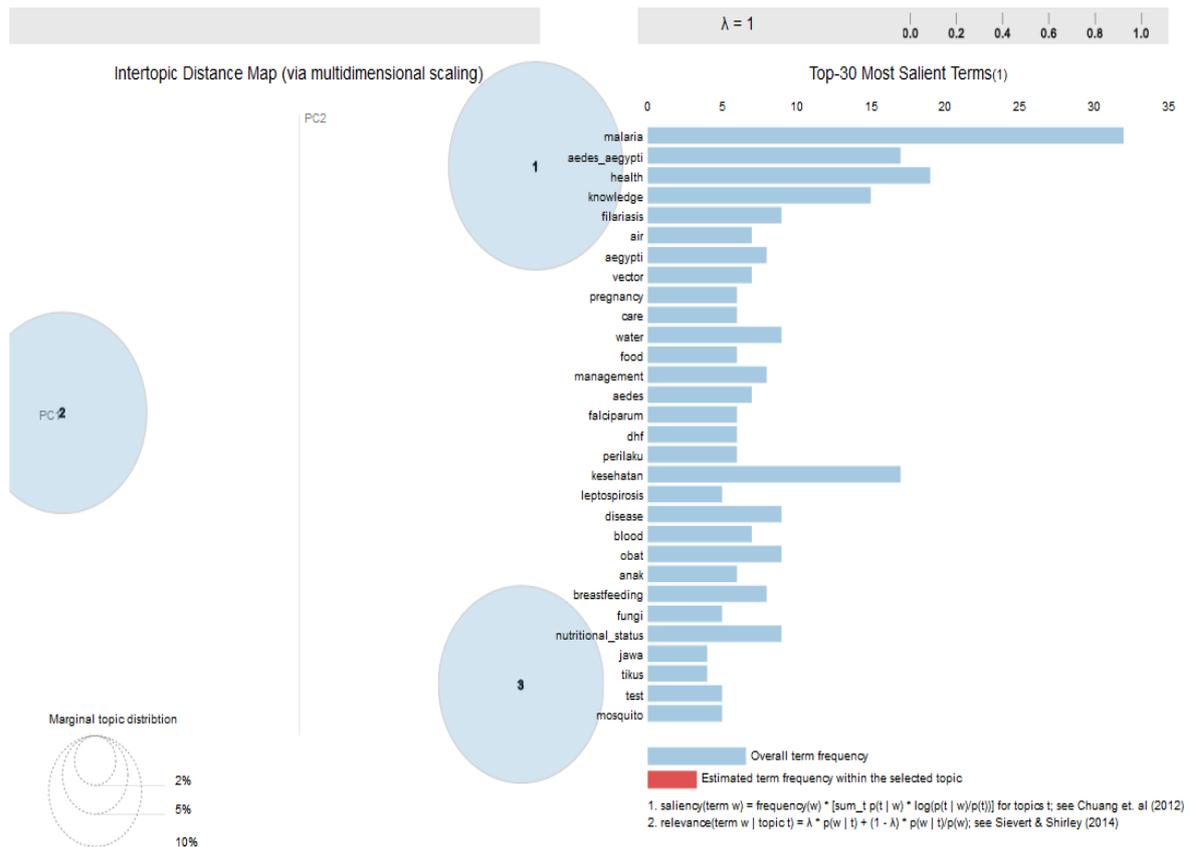
4.4.5 Visualisasi Pemodelan Topik

Setelah selesai melakukan pemodelan topik, model tersebut disimpan ke dalam bentuk *pyLDAvis* dan *word cloud*. Kedua model visualisasi ini dianggap cukup representatif untuk menampilkan hasil pemodelan topik. Untuk *pyLDAvis*, hasilnya disimpan ke dalam file dengan format html. Adapun untuk implementasi kode program dapat dilihat pada Gambar 4.25 dengan tampilan isi file yang dapat dilihat pada Gambar 4.26.

```
import gensim
import pyLDAvis.gensim;pyLDAvis.enable_notebook()

data = pyLDAvis.gensim.prepare(model, corpus_tfidf, dictionary)
print(data)
pyLDAvis.save_html(data, 'lda-gensim-2019.html')
```

Gambar 4.25 Cuplikan kode menyimpan model dalam bentuk *pyLDAvis*



Gambar 4.26 Hasil Visualisasi pyLDAvis

Pada Gambar 4.26 menggambarkan 30 kata penting yang muncul di *corpus*. Panel kanan menggambarkan tentang kata-kata dominan yang dibahas topik dari dataset. Berdasarkan panel kanan visualisasi, penampilan istilah Malaria menjadi kata yang paling sering muncul di *corpus*. Oleh karena itu, peneliti menyimpulkan bahwa penelitian kesehatan di Indonesia terindeks pada SINTA yang paling banyak diteliti yaitu tentang penyakit malaria. Oleh karena itu pemerintah Indonesia seharusnya sangat penting untuk menyebarkan informasi tentang penyakit malaria kepada masyarakat mulai dari pencegahan, penanganan, dan pengobatan.

Sedangkan untuk visualisasi *wordcloud*, disimpan ke dalam file dengan format png. Adapun untuk implementasi kode program dapat dilihat pada Gambar 4.27 dengan contoh hasil keluaran yang dapat dilihat pada Gambar 4.28.

Pada Gambar 4.28 tersebut yaitu untuk mengetahui term judul penelitian kesehatan Indonesia menggunakan visualisasi *wordcloud*. Hasil dari visualisasi *wordcloud* muncul Berbahasa Indonesia dan Bahasa Inggris karena menyesuaikan topik (*keyword*) di jurnal penelitian kesehatan Indonesia yang terindeks SINTA. Dari visualisasi *wordcloud* tersebut kita bisa melihat kata-kata apa yang paling signifikan yang membentuk segmen tertentu dari model yang dibangun. Berikut hasil model topik yang terbentuk didalam *wordcloud* dapat dilihat pada Tabel 4.3.

Tabel 4.3 Hasil Pemodelan Topik Secara Keseluruhan

Topic	Term
Topik #1:	<i>Water, Health, Malaria, Filariasis, Nutritional status, Disease</i>
Topik #2:	<i>Management Health, Air, Risk, Health, Anophles, Tradisional</i>
Topik #3:	<i>Malaria, Aedes aegypti, Anophles, Dengue, Food, Knowledge</i>

4.4.6 Pemodelan Topik Pada Tahun-Tahun Tertentu

Pemodelan ini dilakukan untuk melakukan identifikasi terhadap riset-riset penelitian kesehatan di Indonesia pada tahun-tahun tertentu. Pemodelan ini diperlukan bertujuan untuk melihat gambaran tren penelitian di bidang kesehatan dari tahun 2017 hingga tahun 2019. Dengan tahap-tahapan yang sama berikut hasil topik-topik penelitian kesehatan di Indonesia berdasarkan tahun dapat dilihat pada tabel 4.4, 4.5, 4.6.

Tabel 4.4 Topik-Topik Pada Tahun 2017

Topic	Term
Topik #1:	0.010*"rumah_sakit" + 0.008*"kesehatan" + 0.007*"fasilitas_kesehatan" + 0.006*"kebijakan" + 0.006*"pelaksanaan_program" + 0.006*"bidan_desa" + 0.006*"puskesmas" + 0.006*"faktor_berhubungan" + 0.005*"jurnal" + 0.005*"baru"
Topik#2:	0.016*"jaminan_kesehatan" + 0.015*"studi_kasus" + 0.013*"kesehatan" + 0.013*"kebijakan" + 0.012*"ibu_anak" + 0.011*"pelayanan_kesehatan" + 0.010*"puskesmas" + 0.008*"jaminan" + 0.008*"kabupaten" + 0.008*"dinas_kesehatan"
Topik #3:	0.019*"implementasi_kebijakan" + 0.008*"berat_badan" + 0.005*"health" + 0.005*"jawa_timur" + 0.004*"kematian_bayi" + 0.004*"implementasi" + 0.004*"risk_factors" +

	0.004*"asi_eksklusif" + 0.004*"hubungan" + 0.004*"kawasan_rokok"
--	---

Tabel 4.5 Topik-Topik Pada Tahun 2018

Topic	Term
Topik #1:	0.005*"status_gizi" + 0.005*"factors" + 0.005*"tekanan_darah" + 0.004*"risk_factors" + 0.004*"working_area" + 0.004*"analysis" + 0.004*"hubungan" + 0.003*"health" + 0.003*"rumah_sakit" + 0.003*"hubungan_tingkat"
Topik#2:	0.012*"faktor_faktor" + 0.011*"wilayah_kerja" + 0.008*"faktor" + 0.006*"ibu_hamil" + 0.005*"hiv_aids" + 0.005*"pelayanan_kesehatan" + 0.005*"kesehatan" + 0.005*"puskesmas" + 0.005*"studi_kasus" + 0.004*"ibu"
Topik #3:	0.007*"rumah_sakit" + 0.005*"bandar_lampung" + 0.005*"pasien" + 0.004*"pengaruh_pemberian" + 0.004*"ekstrak_etanol" + 0.004*"public_health" + 0.004*"anak_usia" + 0.004*"pengaruh" + 0.003*"pasien_diabetes" + 0.003*"pemanfaatan_pelayanan"

Tabel 4.6 Topik-Topik Pada Tahun 2019

Topic	Term
Topik #1:	0.009*"status_gizi" + 0.009*"malaria" + 0.007*"remaja_putri" + 0.006*"pendidikan_kesehatan" + 0.005*"hubungan" + 0.005*"studi_kasus" + 0.005*"pengaruh_pemberian" + 0.005*"pengaruh" + 0.005*"pengetahuan" + 0.004*"gizi"
Topik#2:	0.014*"rumah_sakit" + 0.005*"palembang_tahun" + 0.005*"malaria" + 0.005*"sakit" + 0.005*"pengetahuan_sikap" + 0.005*"tahun" + 0.004*"pengaruh_terapi" + 0.004*"pasien" + 0.004*"hubungan" + 0.004*"pemberian_asi"
Topik #3:	0.019*"faktor_faktor" + 0.014*"malaria" + 0.013*"faktor" + 0.007*"berhubungan" + 0.005*"ibu_hamil" + 0.005*"berhubungan_kejadian" + 0.005*"anak_usia" + 0.005*"kerja" + 0.005*"puskesmas" + 0.005*"kejadian"

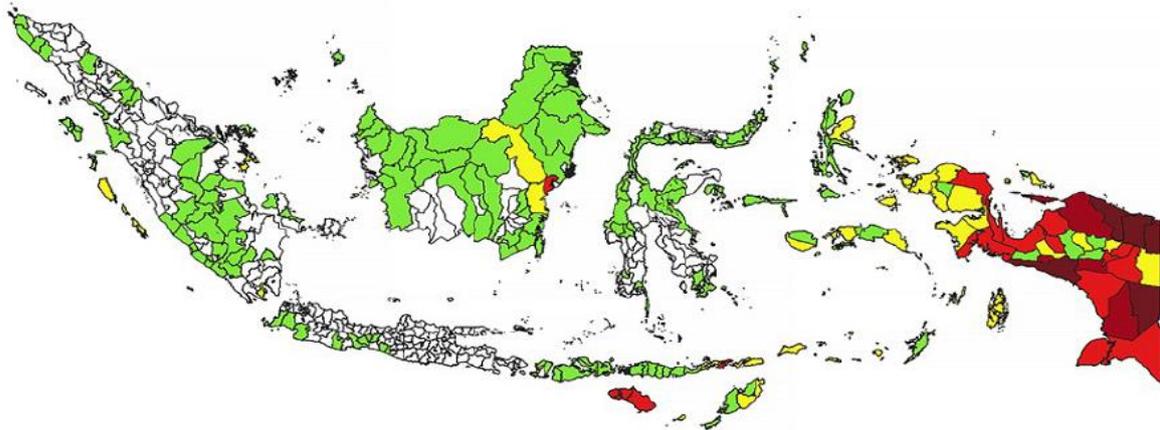
Pada tabel 4.4, 4.5 dan 4.6 hasil pemodelan topik yang dihasilkan pada tahun 2017, 2018, dan 2019. Pada tahun 2017 topik yang tren diteliti yaitu secara garis besar yaitu tentang gizi, malaria, tekanan darah, rumah sakit, jaminan kesehatan, ibu anak, pelayanan kesehatan, dan dinas kesehatan. Pada tahun 2018 topik yang tren diteliti yaitu secara garis besar yaitu terkait implementasi kebijakan, malaria, berat badan, kematian bayi, asi eksklusif, kawasan rokok, ibu hamil, hiv aids, pelayanan kesehatan, dan puskesmas. Pada tahun 2019 topik yang tren diteliti yaitu secara garis besar yaitu terkait status gizi, remaja putri, malaria, pengaruh terapi, pemberian asi, dan ibu hamil.

4.5 Analisis dan Evaluasi

Penelitian ini dapat mengimplementasikan jejaring penelitian kesehatan di Indonesia dengan memanfaatkan *query* yang didapatkan dari pemodelan yang dihasilkan. Penelitian ini data diperoleh dari riset-riset khusus di bidang kesehatan yang terindeks SINTA dengan *keyword* kesehatan dan *health*. Penelitian kesehatan di bidang kesehatan masih banyak yang tidak terindeks selain SINTA, Maka pada penelitian ini hasil pemodelan topik yang dihasilkan belum bisa digunakan sebagai model yang general untuk seluruh Indonesia, karena banyak penelitian-penelitian pada bidang kesehatan yang dipublikasikan di luar SINTA. Metode yang digunakan yaitu dengan membangun pengetahuan dari pemodelan pengetahuan yang dihasilkan dengan *graph database*, Kemudian untuk mengetahui informasi yang bermanfaat domain medis, penelitian ini menerapkan *query* yang dipotong-potong berdasarkan *keyword*, tahun, judul, afiliasi, dan *author*. *Query* yang dapat dilakukan pada penelitian ini bertujuan untuk mengetahui topik-topik penelitian pada tahun tertentu, mengetahui daerah afiliasi pada topik penelitian, mengetahui judul-judul penelitian, dan mengetahui *author* penelitian. Hasil dari *query* kemudian diekstraks untuk mengetahui topik riset-riset di Indonesia yang terindeks SINTA. Hasil Analisis penelitian ini diharapkan dapat membantu Peneliti Kesehatan, Pelayan Kesehatan, Pemerintah dalam mengambil keputusan dalam hal memberikan pelayanan kesehatan di Indonesia.

Dari hasil *query* dianalisis lebih lanjut untuk mengetahui tren topik penelitian di Indonesia yang terindeks SINTA dengan *topic modelling*. Hasil dari tren penelitian-penelitian kesehatan di Indonesia yang terindeks SINTA dengan *keyword* kesehatan dan *health* dari tahun 2017 sampai dengan 2019 yaitu tentang air minum, penyakit payudara, manajemen kesehatan, nyamuk *anophles*, *aedes aegypti*, *dengue*. Dari tren penelitian tersebut bobot yang paling tinggi yaitu malaria. Hal ini terbukti pemerintah Indonesia

memandang penyakit malaria sebagai ancaman terhadap kesehatan masyarakat terutama rakyat yang hidup di daerah terpencil. Dalam Peraturan Presiden No. 2 tahun 2015 dalam RPJMN (Rencana Pembangunan Jangka Menengah Nasional) tahun 2018-2024 dimana malaria termasuk penyakit prioritas yang perlu ditanggulangi. Malaria termasuk penyakit prioritas yang perlu ditanggulangi. Menurut WHO dan Kemenkes RI (Com 2017) peta persebaran malaria di Indonesia disajikan pada Gambar 4.29 sebagai berikut:



Gambar 4.29 Peta Persebaran Penyakit Malaria Indonesia

Pada Gambar 4.29. 70 juta orang Indonesia berada di wilayah endemik malaria, 56,3 tinggal di daerah endemik malaria sedang, dengan 15 juta kasus malaria klinis. Dari 293 kota yang ada di Indonesia, 167 kota endemis malaria dan lebih dari 422 macam spesies nyamuknya. Hampir separoh warga Indonesia tinggal di wilayah yang endemik malaria dan ditemukan 30 juta kasus setiap tahunnya. Banyak kejadian disebabkan permasalahan teknis seperti pembangunan yang tidak berwawasan kesehatan lingkungan, mobilitas penduduk dari daerah endemis malaria, adanya resistensi (Kebal) nyamuk vektor terhadap insektisida yang digunakan, dan resistensi terhadap obat malaria yang makin meluas penyakit malaria endemitas sudah menyebar di seluruh provinsi di Indonesia. Di Indonesia, penyakit malaria menjadi penyakit yang paling banyak ditemukan di Papua, Nusa Tenggara Timur, Sulawesi Tengah, dan Maluku. Malaria merupakan salah satu masalah kesehatan masyarakat Indonesia yang dapat menyebabkan kematian pada kelompok resiko yang tinggi seperti bayi, anak balita, ibu hamil. Penyakit malaria merupakan salah satu penyakit menular. Pengendalian malaria dapat dilakukan dengan upaya preventif, promotif, kuratif dan rehabilitatif yang bertujuan untuk menurunkan angka kematian dan kejadian luar biasa (KLB). Untuk mencapai hasil maksimal, berkualitas dan optimal upaya harus dilakukan integrasi dengan layanan kesehatan dan program-program lain. Titik berat

penatalaksanaan penanganan kasus malaria yang optimal dan berkualitas diharapkan akan memberikan kontribusi langsung menuju Indonesia bebas penyakit malaria. Oleh karena itu perlu pemerintah dalam menyebarkan informasi yang berkaitan penyakit malaria mulai dari penanganan, pencegahan dan pengobatan. Penelitian ini tidak menutup kemungkinan dengan proses *topic modelling* yang lainnya sebagai catatan untuk penelitian yang lebih lanjut.

Pengujian dalam penelitian ini menggunakan kuesioner sederhana yang disampaikan responden yang terdiri peneliti, tenaga kesehatan, dan akademis. Pengujian yang diuji hanya berupa pengujian fungsional hasil pemodelan. Pengujian ini menanyakan kesesuaian bagaimana pendapat terhadap pemodelan basis pengetahuan jejaring penelitian kesehatan yang dihasilkan saja. Pengujian dilakukan sebanyak 50 responden yang terdiri dari perawat, dokter, peneliti dan mahasiswa. Hasil penilaian pengujian responden dan instrumen yang dibangun dapat dilihat pada Tabel 4.7.

Tabel 4.7 Hasil Pengujian Kuesioner

No	Pertanyaan	Hasil			
		SB	B	C	K
1	Apakah pemodelan pengetahuan Jejaring penelitian dengan menggunakan <i>text mining</i> yang dibangun mudah dipahami?	43	7	0	0
2	Apakah pemodelan pengetahuan Jejaring penelitian dengan menggunakan <i>text mining</i> memudahkan para pengguna melakukan analisis penelitian kesehatan di Indonesia?	46	4	0	0
3	Apakah pemodelan pengetahuan Jejaring penelitian dengan menggunakan <i>text mining</i> yang dihasilkan sudah sesuai?	43	4	3	0
4	Apakah pemodelan pengetahuan Jejaring penelitian dengan menggunakan <i>text mining</i> bermanfaat bagi peneliti, tenaga kesehatan, dan akademis?	45	4	1	0
5	Apakah pemodelan yang disediakan dapat mampu menunjang untuk mengembangkan penelitian kesehatan di Indonesia?	47	3	0	0

Total Pengujian Responden	228	18	4	0
----------------------------------	-----	----	---	---

Keterangan:

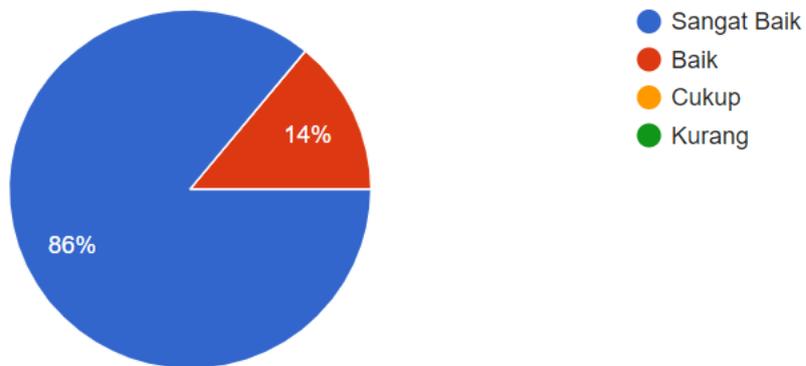
SB : Sangat Baik

B : Baik

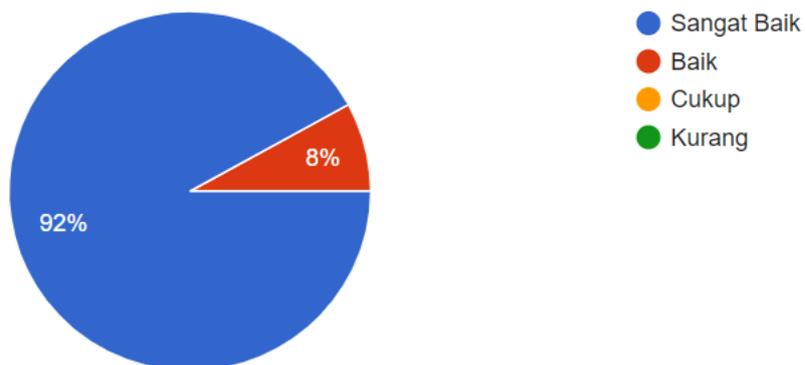
C : Cukup

K : Kurang

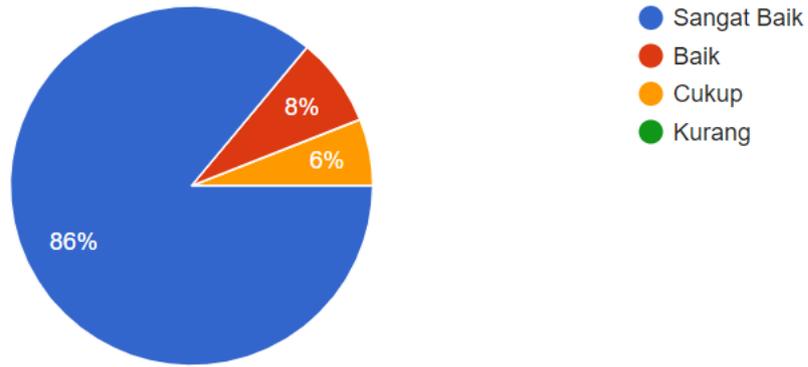
Adapun Gambar persentase masing kuesioner dapat dilihat pada Gambar 4.30, 4.31, 4.32, 4.33 dan 4.34.



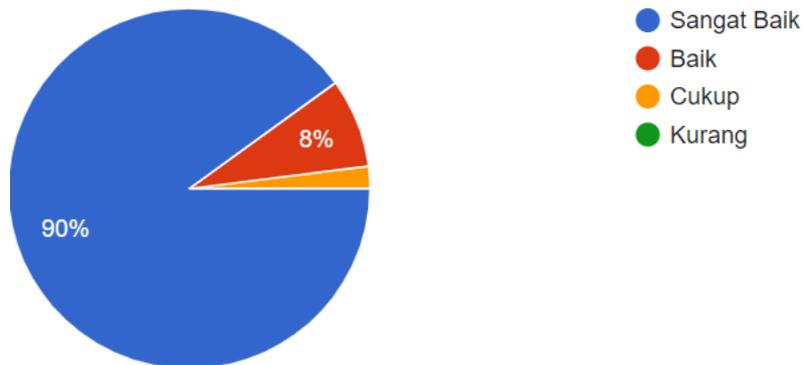
Gambar 4.30 Persentase Pertanyaan kuisoner nomor 1



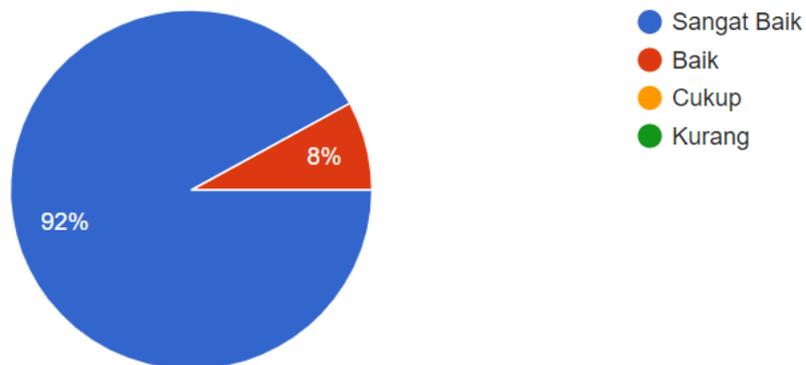
Gambar 4.31 Persentase Pertanyaan kuisoner nomor 2



Gambar 4.32 Persentase Pertanyaan kuisioner nomor 3



Gambar 4.33 Persentase Pertanyaan kuisioner nomor 4



Gambar 4.32 Persentase Pertanyaan kuisioner nomor 5

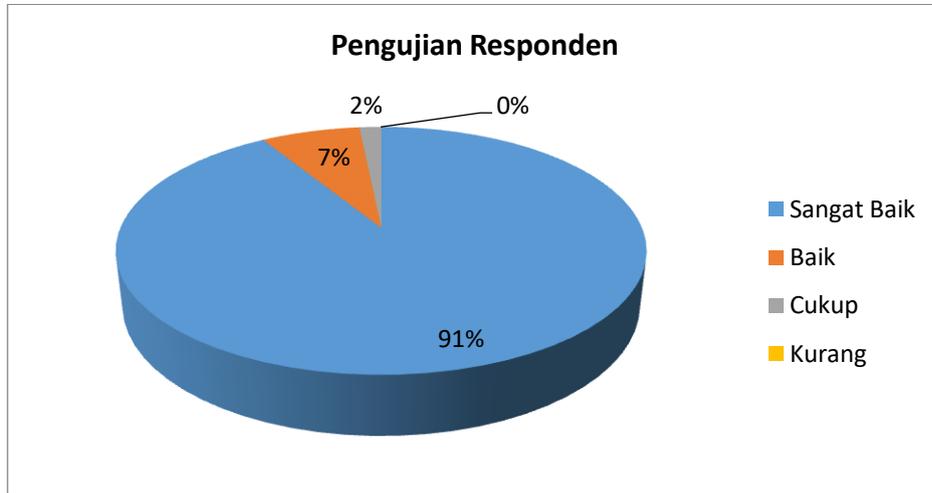
Nilai total secara keseluruhan pengujian kuesioner adalah 250. Responden yang melakukan pengujian memberikan tanggapan sangat baik sebanyak 91,2%, baik 7,2%, cukup 1,6%. Adapun perhitungan pengujian sebagai berikut:

$$\text{Sangat Baik} = \frac{228}{250} \times 100\% = 91,2\%$$

$$\text{Baik} = \frac{18}{250} \times 100\% = 7,2\%$$

$$Cukup = \frac{4}{250} \times 100\% = 1,6\%$$

Adapun persentase disajikan pada Gambar 4.35.



Gambar 4.35 Hasil Pengujian Responden

Evaluasi terhadap penelitian ini tidak menutup kemungkinan dengan proses penurunan *query-query* lainnya yang dapat dikembangkan untuk penelitian lebih lanjut sesuai dengan keperluan peneliti.

BAB 5

Kesimpulan dan Saran

5.1 Kesimpulan

Penelitian berhasil mengimplementasikan jejaring penelitian kesehatan di Indonesia dengan memanfaatkan *query* yang di dapatkan dari pemodelan yang dihasilkan. Data yang dihasilkan dari SINTA dengan *keyword* kesehatan dan *health*. Metode yang digunakan yaitu dengan membangun pengetahuan dari pemodelan pengetahuan yang dihasilkan dengan *graph database*, kemudian untuk mengetahui informasi yang bermanfaat domain medis dengan menerapkan *query* yang dipotong-potong berdasarkan *keyword*, tahun, judul, afiliasi, dan *author*. *Query* yang dapat dilakukan pada penelitian ini dapat mengetahui topik-topik penelitian pada tahun tertentu, mengetahui daerah afiliasi pada topik penelitian, mengetahui judul-judul penelitian, mengetahui *author* penelitian. Hasil dari *query* kemudian di ekstrak untuk mengetahui topik riset-riset di Indonesia.

Hasil dari penelitian ini berhasil melakukan pemodelan topik, model tersebut disimpan ke dalam bentuk *pyLDavis* dan *word cloud*. Hasil visualisasi hasil pemodelan topik, *word cloud* dan *pyLDavis* dapat digunakan untuk memberikan gambaran persebaran kata terhadap topik dan relevansinya terhadap topik tersebut. Kedua model visualisasi ini dianggap cukup representatif untuk menampilkan hasil pemodelan topik. Adapun hasil dari visualisasi *pyLDavis*, hasilnya disimpan ke dalam file dengan format html. Untuk visualisasi *wordcloud* tersebut kita bisa melihat kata-kata apa yang paling signifikan yang membentuk segmen tertentu dari model yang dibangun. Pemodelan topik yang dihasilkan tren penelitian yang kesehatan di Indonesia yaitu tentang air minum, penyakit payudara, manajemen kesehatan, nyamuk anophles, aedes aegypti, dengue. Dari tren penelitian tersebut bobot yang paling tinggi yaitu malaria. Hasil Analisis penelitian ini diharapkan dapat membantu Peneliti Kesehatan, Pelayanan Kesehatan, Pemerintah dalam mengambil keputusan dalam hal memberikan pelayanan kesehatan di Indonesia.

5.2 Saran

Saran untuk pengembangan sistem ini maupun penelitian yang selanjutnya adalah:

- a. Dapat mengeksplor kembali *query-query* yang dihasilkan untuk bermanfaat dalam domain medis dan dapat dikembangkan bagi peneliti dalam penelitian yang dilakukan khususnya menggunakan *graph* database Neo4j sesuai dengan bidang yang diteliti.
- b. Penerapan metode membangun pengetahuan yang berbeda domain selain medis untuk analisis jejaring dalam bidang lainnya menggunakan *graph* database Neo4j.
- c. Implementasi topik pemodelan berbagai skenario yang berbeda, baik dalam tahap preprocessing, nilai parameter untuk setiap fungsi, penentuan nilai koherensi, dan lain sebagainya untuk menguji dan mengevaluasi lebih dalam bagaimana performa metode *Latent Dirichlet Allocation* (LDA) untuk melakukan pemodelan topik.

Daftar Pustaka

- Bajaj, Vasvi, Rajat Bhusan Panda, Chetna Dabas, and Parmeet Kaur. 2018. "Graph Database for Recipe Recommendations." *2018 7th International Conference on Reliability, Infocom Technologies and Optimization: Trends and Future Directions, ICRITO 2018*, 276–81. <https://doi.org/10.1109/ICRITO.2018.8748827>.
- Banazir, B, and Annes Philip. 2013. "Efficient Keyword Search Using Text Mining Techniques : A Survey" 3 (1): 473–75.
- Barat, Jawa, Ekstrak Propolis, Memperbaiki Profil, Berat Badan, Tikus Model, Penentuan Nyamuk, and Tenggara Timur. n.d. "MEDIA PENELITIAN DAN."
- Bratawisnu, Kevin Made, and Andry Alamsyah. 2018. "Social Network Analysis Untuk Analisa Interaksi User Dimedia Sosial Mengenai Bisnis E-Commerce (Studi Kasus: Lazada, Tokopedia Dan Elevenia)." *Jurnal Manajemen Dan Bisnis (Almana) Vol. 2* (2): 107–15. <http://journal.unla.ac.id/index.php/almana/article/view/143>.
- Buenano-Fernandez, Diego, Mario Gonzalez, David Gil, and Sergio Lujan-Mora. 2020. "Text Mining of Open-Ended Questions in Self-Assessment of University Teachers: An LDA Topic Modeling Approach." *IEEE Access* 8 (February): 35318–30. <https://doi.org/10.1109/ACCESS.2020.2974983>.
- Com, Iv. 2017. *STATE OF HEALTH INEQUALITY Indonesia*. <http://apps>.
- Constantinov, Calin, Lucian Iordache, Adrian Georgescu, Paul Stefan Popescu, and Mihai Mocanu. 2018. "Performing Social Data Analysis with Neo4j: Workforce Trends and Corporate Information Leakage." *2018 22nd International Conference on System Theory, Control and Computing, ICSTCC 2018 - Proceedings*, 403–6. <https://doi.org/10.1109/ICSTCC.2018.8540645>.
- Dharmawan, I. Nyoman Pande Wahyu, and Riyanarto Sarno. 2017. "Book Recommendation Using Neo4j Graph Database in BibTeX Book Metadata." *Proceeding - 2017 3rd International Conference on Science in Information Technology: Theory and Application of IT for Education, Industry and Society in Big Data Era, ICSITech 2017* 2018-Janua: 47–52. <https://doi.org/10.1109/ICSITech.2017.8257084>.
- Dunkel, Alexander, Gennady Andrienko, Natalia Andrienko, Dirk Burghardt, Eva Hauthal, and Ross Purves. 2019. "A Conceptual Framework for Studying Collective Reactions to Events in Location-Based Social Media." *International Journal of Geographical Information Science* 33 (4): 780–804.

- <https://doi.org/10.1080/13658816.2018.1546390>.
- Feldman, Ronen, and James Sanger. 2006. *The Text Mining Handbook. The Text Mining Handbook*. <https://doi.org/10.1017/cbo9780511546914>.
- Francis, Nadime, Alastair Green, Paolo Guagliardo, Leonid Libkin, Tobias Lindaaker, Victor Marsault, Stefan Plantikow, Mats Rydberg, Petra Selmer, and Andrés Taylor. 2018. “Cypher: An Evolving Query Language for Property Graphs.” *Proceedings of the ACM SIGMOD International Conference on Management of Data*, 1433–45. <https://doi.org/10.1145/3183713.3190657>.
- Jo, Taeho. 2018. *(Studies in Big Data) Taeho Jo-Text Mining_ Concepts, Implementation, and Big Data Challenge-Springer (2018).Pdf*.
- Johnpaul, C. I., and Tojo Mathew. 2017. “A Cypher Query Based NoSQL Data Mining on Protein Datasets Using Neo4j Graph Database.” *2017 4th International Conference on Advanced Computing and Communication Systems, ICACCS 2017*, 4–9. <https://doi.org/10.1109/ICACCS.2017.8014558>.
- Kusumaningrum, Retno, M. Ihsan Aji Wiedjayanto, Satriyo Adhy, and Suryono. 2017. “Classification of Indonesian News Articles Based on Latent Dirichlet Allocation.” *Proceedings of 2016 International Conference on Data and Software Engineering, ICoDSE 2016*, 1–5. <https://doi.org/10.1109/ICODSE.2016.7936106>.
- Lakshmi Prasanna, P., and D. Rajeswara Rao. 2019. “A Text Mining Research Based on Topic Modeling Using Latent Dirichlet Allocation.” *International Journal of Recent Technology and Engineering* 7 (5): 308–17.
- Lu, Huiling, Zhiguo Hong, and Minyong Shi. 2017. “Analysis of Film Data Based on Neo4j.” *Proceedings - 16th IEEE/ACIS International Conference on Computer and Information Science, ICIS 2017*, 675–77. <https://doi.org/10.1109/ICIS.2017.7960078>.
- Ma, Yisong, Zhigang Wu, Lin Guan, Baorong Zhou, and Rongrong Li. 2014. “Study on the Relationship between Transmission Line Failure Rate and Lightning Information Based on Neo4j.” *POWERCON 2014 - 2014 International Conference on Power System Technology: Towards Green, Efficient and Smart Power System, Proceedings*, no. Powercon: 474–79. <https://doi.org/10.1109/POWERCON.2014.6993713>.
- Ministry of Health Republic of Indonesia. 2016. *2015 Indonesia Health Profile*.
- O'Malley, A. James, and Peter V. Marsden. 2008. “The Analysis of Social Networks.” *Health Services and Outcomes Research Methodology* 8 (4): 222–69. <https://doi.org/10.1007/s10742-008-0041-z>.
- Prihatini, P. M., Ikgd Putra, Iad Giriantari, and M. Sudarma. 2017. “Indonesian Text

- Feature Extraction Using Gibbs Sampling and Mean Variational Inference Latent Dirichlet Allocation.” *QiR 2017 - 2017 15th International Conference on Quality in Research (QiR): International Symposium on Electrical and Computer Engineering* 2017-Decem: 40–44. <https://doi.org/10.1109/QIR.2017.8168448>.
- Rabuzin, Kornelije, and Martina Sestak. 2018. “Towards Inheritance in Graph Databases.” *2018 International Conference on Information Management and Processing, ICIMP 2018* 2018-Janua: 115–19. <https://doi.org/10.1109/ICIMP1.2018.8325851>.
- Rahmi Surayya. 2018. “Pendekatan Kualitatif Dalam Penelitian Kesehatan.” *Jurnal Kedokteran Dan Kesehatan Malikussaleh*, 75–84.
- Sistem, Rekayasa, Yoga Sahria, and Dthomas Hatta Fudholi. 2021. “JURNAL RESTI Analisis Topik Penelitian Kesehatan Di Indonesia Menggunakan Metode” 1 (10): 336–44.
- Sri Arini, Ni Wayan, Ida Bagus Putu Widja, and I Komang Rinartha Yasa Negara. 2019. “Analisis Frekuensi Kata Untuk Mengekstrak Kata Kunci Dari Artikel Ilmiah Berbahasa Indonesia.” *Eksplora Informatika* 8 (2): 80–84. <https://doi.org/10.30864/eksplora.v8i2.162>.
- Sutrisno, Tri. 2017. “Aplikasi Graf Dalam Rekayasa Perangkat Lunak.” *Jurnal Muara Sains, Teknologi, Kedokteran Dan Ilmu Kesehatan* 1 (1): 318–27. <https://doi.org/10.24912/jmstkik.v1i1.445>.
- Tong, Zhou, and Haiyi Zhang. 2016. “A Text Mining Research Based on LDA Topic Modelling,” 201–10. <https://doi.org/10.5121/csit.2016.60616>.
- Wardhani, Yurika Fauzia, and Astridya Paramita. 2016. “Pelayanan Kesehatan Mental Dalam Hubungannya Dengan Disabilitas Dan Gaya Hidup Masyarakat Indonesia (Analisis Lanjut Riskesdas 2007 Dan 2013).” *Buletin Penelitian Sistem Kesehatan* 19 (1): 99–107.
- Wirawan, Panji Wisnu, Djalal Er Riyanto, Dinar Mutiara Kusumo Nugraheni, and Yasmin Yasmin. 2019. “Graph Database Schema for Multimodal Transportation in Semarang.” *Journal of Information Systems Engineering and Business Intelligence* 5 (2): 163. <https://doi.org/10.20473/jisebi.5.2.163-170>.
- Wirawan, Panji Wisnu, and Djalal Er Riyanto. 2017. “Kajian Implementasi Graph Database Pada Rute Bus Rapid Transit.” *Jurnal Nasional Teknologi Dan Sistem Informasi* 3 (3): 313–19. <https://doi.org/10.25077/teknosi.v3i3.2017.313-319>.
- Wirawan, Panji Wisnu, Djalal Er Riyanto, and Khadijah Khadijah. 2016. “Pemodelan Graph Database Untuk Moda Transportasi Bus Rapid Transit.” *Jurnal Informatika* 10

- (2): 1271–79. <https://doi.org/10.26555/jifo.v10i2.a5072>.
- Wita, Ratsameetip, Kawinwit Bubphachuen, and Jakarin Chawachat. 2018. “Content-Based Filtering Recommendation in Abstract Search Using Neo4j.” *ICSEC 2017 - 21st International Computer Science and Engineering Conference 2017, Proceeding* 6: 136–39. <https://doi.org/10.1109/ICSEC.2017.8443957>.
- Yudiarta, Nyoman Gede, Made Sudarma, and Wayan Gede Ariastina. 2018. “Penerapan Metode Clustering Text Mining Untuk Pengelompokan Berita Pada Unstructured Textual Data.” *Majalah Ilmiah Teknologi Elektro* 17 (3): 339. <https://doi.org/10.24843/mite.2018.v17i03.p06>.
- Yusainy, Cleoputri, Anif Fatma Chawa, and Siti Kholifah. 2017. “Social Data Analytics Sebagai Metode Alternatif Dalam Riset Psikologi.” *Buletin Psikologi* 25 (2): 67–75. <https://doi.org/10.22146/buletinpsikologi.27751>.
- Zhao, Jian, Zhiguo Hong, and Minyong Shi. 2019. “Analysis of Disease Data Based on Neo4J Graph Database.” *Proceedings - 18th IEEE/ACIS International Conference on Computer and Information Science, ICIS 2019*, 381–84. <https://doi.org/10.1109/ICIS46139.2019.8940247>.

LAMPIRAN A

1. Daftar Nama Jurnal Kesehatan di Indonesia

Nomor	Judul
1	Jurnal Ekologi Kesehatan
2	KEMAS : Jurnal Kesehatan Masyarakat
3	Media Penelitian dan Pengembangan Kesehatan
4	Jurnal Kesehatan Lingkungan Indonesia
5	Medisains : Jurnal Ilmiah Ilmu - Ilmu Kesehatan
6	Jurnal Administrasi Kesehatan Indonesia
7	Media Kesehatan Masyarakat Indonesia
8	Buletin Penelitian Kesehatan
9	Buletin Penelitian Sistem Kesehatan
10	Berkala Ilmu Kesehatan Kulit dan Kelamin (Periodical of Dermatology and Venerology)
11	Jurnal Kesehatan Lingkungan
12	JKKI : Jurnal Kedokteran dan Kesehatan Indonesia
13	Jurnal Kesehatan Reproduksi
14	Jurnal Endurance: Kajian Ilmiah Problema Kesehatan
15	Jurnal Vokasi Kesehatan
16	Jurnal Ilmu Kesehatan Masyarakat : The Public Health Science Journal
17	Jurnal Kebijakan Kesehatan Indonesia : JKKI
18	Jurnal Kesehatan Prima
19	Jurnal Epidemiologi Kesehatan Komunitas
20	Jurnal Kesehatan Manarang
21	Jurnal Info Kesehatan
22	Jurnal Kesehatan Andalas
23	Jurnal Ilmiah Ibnu Sina (JIIS): Ilmu Farmasi dan Kesehatan
24	STRADA Jurnal Ilmiah Kesehatan
25	Mutiara Medika: Jurnal Kedokteran dan Kesehatan
26	Jurnal Promosi Kesehatan Indonesia
27	Jurnal Ilmu Kesehatan Masyarakat
28	Jurnal Kesehatan
29	Window of Health : Jurnal Kesehatan

30	BioLink (Jurnal Biologi Lingkungan, Industri, Kesehatan)
31	Riset Informasi Kesehatan
32	Jurnal Ekonomi Kesehatan Indonesia
33	Dunia Keperawatan : Jurnal Keperawatan dan Kesehatan
34	Kes Mas: Jurnal Fakultas Kesehatan Masyarakat
35	Jurnal Kesehatan Reproduksi
36	Jurnal Kesehatan Gigi
37	Majalah Kesehatan
38	Jurnal Kesehatan Lingkungan: Jurnal dan Aplikasi Teknik Kesehatan Lingkungan
39	Jurnal Penelitian Kesehatan Suara Forikes
40	Jurnal Profesi Medika : Jurnal Kedokteran dan Kesehatan
41	Jurnal Kesehatan Ibu dan Anak
42	Jurnal Kesehatan Masyarakat Andalas (Andalas Journal of Public Health)
43	GASTER: Jurnal Kesehatan
44	Jurnal Ilmu dan Teknologi Kesehatan
45	SCIENTIA Jurnal Farmasi dan Kesehatan
46	MEDIA ILMU KESEHATAN
47	Media Karya Kesehatan
48	JUMANTIK (Jurnal Ilmiah Penelitian Kesehatan)
49	Jurnal Kesehatan Vokasional
50	Jurnal Kesehatan Lingkungan
51	Care : Jurnal Ilmiah Ilmu Kesehatan
52	Jurnal Manajemen Kesehatan Yayasan RS.Dr.Soetomo
53	Jurnal Kesehatan Bakti Tunas Husada: Jurnal Ilmu-ilmu Keperawatan, Analisis Kesehatan dan Farmasi
54	Jp.jok (Jurnal Pendidikan Jasmani, Olahraga dan Kesehatan)
55	Jurnal Sains dan Kesehatan
56	Sanitasi: Jurnal Kesehatan Lingkungan
57	Jurnal Manajemen Informasi Kesehatan Indonesia (JMiki)
58	MPPKI (Media Publikasi Promosi Kesehatan Indonesia): The Indonesian Journal of Health Promotion
59	Jurnal Kesehatan Kusuma Husada
60	Holistik Jurnal Kesehatan

61	Saintika Medika : Jurnal Ilmu Kesehatan dan Kedokteran Keluarga
62	Jurnal Analisis Kesehatan
63	Husada Mahakam: Jurnal Kesehatan
64	Medika Respati : Jurnal Ilmiah Kesehatan
65	Jurnal Aisyah : Jurnal Ilmu Kesehatan
66	Sains Medika : Jurnal Kedokteran dan Kesehatan
67	Jurnal Kesehatan Indonesia (The Indonesian Journal of Health)
68	Soepra Jurnal Hukum Kesehatan
69	Jurnal Kesehatan
70	Jurnal Media Analisis Kesehatan
71	Jurnal Ilmiah Kesehatan
72	Jurnal Kesehatan Mesencephalon
73	Wawasan Kesehatan : Jurnal Ilmiah Ilmu Kesehatan
74	Jurnal Keperawatan dan Kesehatan Masyarakat Cendekia Utama
75	Jurnal Ilmu Kesehatan
76	Jurnal Muara Sains, Teknologi, Kedokteran, dan Ilmu Kesehatan
77	Jurnal Berkala Kesehatan
78	Jurnal Kesehatan
79	Poltekita: Jurnal Ilmu Kesehatan
80	Arkesmas (Arsip Kesehatan Masyarakat)
81	JIK (Jurnal Ilmu Kesehatan)
82	SANITAS : Jurnal Teknologi dan Seni Kesehatan
83	Jurnal Ilkes (Jurnal Ilmu Kesehatan)
84	JPP (Jurnal Kesehatan Poltekkes Palembang)
85	Jurnal Kesehatan Metro Sai Wawai
86	Jurnal Promkes : Jurnal Promosi dan Pendidikan Kesehatan Indonesia
87	Jurnal Wiyata: Penelitian Sains dan Kesehatan
88	Jurnal Kedokteran dan Kesehatan Universitas Sriwijaya
89	VISIKES : Jurnal Kesehatan Masyarakat
90	Berkala Kedokteran : Jurnal Kedokteran dan Kesehatan
91	Jurnal Kesehatan Masyarakat Indonesia
92	Jurnal Kesehatan Komunitas (Journal of Community Health)
93	Jurnal Ilmiah Kesehatan Keperawatan

94	Jurnal Riset Kesehatan Poltekkes Depkes Bandung
95	Jurnal Pendidikan Kesehatan Rekreasi
96	Jurnal Kesehatan
97	Jurnal Kedokteran dan Kesehatan
98	Media Kesehatan Politeknik Kesehatan Makassar
99	Jurnal Kesmas (Kesehatan Masyarakat) Khatulistiwa
100	JKM (Jurnal Kesehatan Masyarakat) Cendekia Utama
101	Jurnal Ilmu Kesehatan
102	Jurnal Manajemen Kesehatan Indonesia
103	Jurnal Kesehatan
104	Quality : Jurnal Kesehatan
105	Jurnal Ilmiah Kesehatan
106	Bravo's : Jurnal Program Studi Pendidikan Jasmani dan Kesehatan
107	Afiasi : Jurnal Kesehatan Masyarakat
108	JURNAL RISET KESEHATAN
109	Hearty : Jurnal Kesehatan Masyarakat
110	Syifa Medika: Jurnal Kedokteran dan Kesehatan
111	Herb-Medicine Journal: Terbitan Berkala Ilmiah Herbal, Kedokteran dan Kesehatan
112	Jurnal Rekam Medis dan Informasi Kesehatan
113	Media Keperawatan: Politeknik Kesehatan Makassar
114	Jurnal Kesehatan
115	Jurnal Mutiara Kesehatan Masyarakat
116	Jurnal Kesehatan Saelmakers PERDANA (JKSP)
117	Majalah Kesehatan Pharmamedika
118	Jurnal Kesehatan Poltekkes Kemenkes RI Pangkalpinang
119	J-KESMAS: Jurnal Kesehatan Masyarakat
120	Jurnal Medikes (Media Informasi Kesehatan)
121	Jurnal Kesehatan Perintis (Perintis's Health Journal)
122	Jurnal Kesehatan Karya Husada
123	Hospital Majapahit : Jurnal Ilmiah Kesehatan Politeknik Kesehatan Mojokerto
124	Jurnal Kesehatan Terpadu
125	Jurnal Kedokteran: Media Informasi Ilmu Kedokteran dan Kesehatan
126	Jambi Medical Journal : Jurnal Kedokteran dan Kesehatan

127	Jurnal Kesehatan Indra Husada
128	Dinamika Kesehatan: Jurnal Kebidanan dan Keperawatan
129	Motorik Jurnal Ilmu Kesehatan
130	Jurnal Bidang Ilmu Kesehatan
131	JIKP Jurnal Ilmiah Kesehatan PENCERAH
132	Journal Of Health Science (Jurnal Ilmu Kesehatan)
133	Wirara Medika: Jurnal Kesehatan
134	Jurnal Publikasi Kesehatan Masyarakat Indonesia
135	Jurnal Kesehatan Medika Sainatika
136	Jurnal Informasi Kesehatan Indonesia (JIKI)
137	Jurnal Kesehatan Masyarakat
138	Jurnal Kesehatan dr. Soebandi
139	JKP (Jurnal Kesehatan Primer)
140	Jurnal Penjakora Fakultas Olahraga dan Kesehatan
141	Jurnal Teknologi Kesehatan (Journal of Health Technology)
142	Jurnal Ilmiah Kesehatan
143	Jurnal Bahana Kesehatan Masyarakat (Bahana of Journal Public Health)
144	Jurnal Kesehatan Masyarakat dan Lingkungan Hidup
145	Buletin Keslingmas : Buletin Kesehatan Lingkungan Masyarakat
146	Bhamada: Jurnal Ilmu dan Teknologi Kesehatan (E-Journal)
147	Jurnal Kesehatan Holistic
148	Jurnal Ilmu dan Teknologi Kesehatan
149	Jurnal Kesehatan
150	MIKIA: Mimbar Ilmiah Kesehatan Ibu dan Anak (Maternal and Neonatal Health Journal)
151	Jurnal Kesehatan Bina Generasi
152	Interest : Jurnal Ilmu Kesehatan
153	Jurnal Kesehatan Medika Udayana
154	Jurnal Kesehatan Terpadu
155	Jurnal Kesehatan Al-Irsyad
156	Jurnal Media Kesehatan
157	Jurnal Penelitian Kesehatan
158	Media Kesehatan Gigi : Politeknik Kesehatan Makassar

159	Jurnal Kejaora (Kesehatan Jasmani dan Olah Raga)
160	Altius: Jurnal Ilmu Olahraga dan Kesehatan
161	Jurnal Wacana Kesehatan
162	Jurnal Ilmiah Kesehatan Ar-Rum Salatiga
163	Jurnal Ilmiah Kesehatan Sandi Husada
164	Surya Medika: Jurnal Ilmiah Ilmu Keperawatan dan Ilmu Kesehatan Masyarakat
165	2-TRIK : TUNAS-TUNAS RISET KESEHATAN
166	Jurnal Pendidikan Kesehatan
167	Medika Kartika : Jurnal Kedokteran dan Kesehatan
168	J-Kesmas: Jurnal Fakultas Kesehatan Masyarakat (The Indonesian Journal of Public Health)
169	Jurnal Penelitian dan Kajian Ilmiah Kesehatan Politeknik Medica Farma Husada Mataram
170	Jurnal Medika : Karya Ilmiah Kesehatan
171	Jakiyah : Jurnal Ilmiah Umum dan Kesehatan Aisyiyah
172	Jurnal Medika : Karya Ilmiah Kesehatan

2. Daftar Responden

No	Nama	Asal
1	dr. Wahyudi Santoso	UNS
2	Endriyatno Sulastomo, S.Kep., M.Kes.	STIKES Wira Husada
3	Nurul Isnaini F, S.Kep., MPH	Rs. Respira Yogyakarta
4	Pamulatsih Dwi O, S.Kep, Ns, M.Kep	Universitas Gadjah Mada
5	Rosikhah Al Maris, S.Kep., Ns., M.Kep	Universitas Gadjah Mada
6	Heni Indarwati, S.Kep, Ns., MMedEd.	Universitas Gadjah Mada
7	Husnul Khotimah, S.Kep,Ns.,MPH	STIKES Al Islam Yogyakarta
8	Sinta Sundari, SIP., S.Kep.	STIKES Al Islam Yogyakarta
9	Candra., MPH	STIKES Al Islam Yogyakarta
10	Edwina Nurutami, M.Kes	STIKES Al Islam Yogyakarta
11	Mirandha Adhihimawati,S.Si.,Msc	Akafarama Yogyakarta
12	Rini Setyowati,S.Farm.,Apt.,M.Farm	Akafarma Yogyakarta
13	Eni Kartikasari,S.Si.,Msc	Akafarama Yogyakarta
14	Sholihatil Hidayati, M.Farm.,Apt	Universitas Brawijaya Malang
15	Kintel Edy Octavianugrah.,M.P.H	STIKES Al Islam Yogyakarta
16	Sigit Prasetyo, M.Pd.	AKPN Bahtera Yogyakarta
17	dr. Ratna	STIKES Al Islam Yogyakarta
18	Wahlah., S.Kep	Ners STIKES Yogyakarta
19	Aini Nurha	Ners STIKES Yogyakarta
20	Anggita Septiana	STIKES Al Islam Yogyakarta
21	Dyah Ayu Surya Wulandari	STIKES Al Islam Yogyakarta
22	Eni Tri Wahyuni	STIKES Al Islam Yogyakarta
23	Muchammad Setiawan	STIKES Al Islam Yogyakarta
24	Anita Sari	STIKES Al Islam Yogyakarta
25	Ayu Pawestri	STIKES Al Islam Yogyakarta
26	Diva Carissima Aisyahra	STIKES Al Islam Yogyakarta
27	Iis Rismagi	STIKES Al Islam Yogyakarta
28	Puspa Dias Vira Ngesti	STIKES Al Islam Yogyakarta
29	Riesky Ardian	STIKES Al Islam Yogyakarta
30	Ririn Dwi Adriani	STIKES Al Islam Yogyakarta

31	Rizal Arianto	STIKES Al Islam Yogyakarta
32	Siti Hamidah	STIKES Al Islam Yogyakarta
33	Fatimah	STIKES Al Islam Yogyakarta
34	Ani Muflikhah	STIKES Al Islam Yogyakarta
35	Indah Septiani	Pukesmas Wonosari
36	Wulan Ramadhan	Rs. JIH
37	Zeni Nur Hikmah Amd.Kep	Rs. Bhayangkara Semarang
38	Dewi Bagaswati, Amd.Kep	Klinik Gunungkidul
39	Miftachul Rohmah, Amd.Kep	Rs. Queen Latifah
40	Nancy Claudia Depi A, Amd.Kep	Rs. Nur Hidayah Yogyakarta
41	Rizky Novita W	Universitas Aisyiyah Surakarta
42	Umi Choiriyah	Universitas Negeri Semarang
43	Olince Yogibalom	STIKES Al Islam Yogyakarta
44	Wiwit Sukmandari	STIKES Al Islam Yogyakarta
45	Ambar Budi Nugroho	STIKES Yogyakarta
46	Bella Puspitasari, S.Kep	RSU. Habibulloh Gabus
47	Lina Puspitasari, Amd.Keb	STIKES Karya Husada Semarang
48	Nurhasyiah	Rumah Sakit Keluarga Sehat
49	Mita Desiriana Sriwulandari	Poltekkes Semarang
50	Danang Yulianto, S.Si., Apt., M.Farm	Akafarma Al Islam Yogyakarta

3. Form Instrumen Kuisioner Pengujian



UNIVERSITAS
ISLAM
INDONESIA

QUISONER PENGUJIAN

Indonesian Medical Research Network Analysis
Through Text Mining

Yoga Sahria
Magister Informatika

Quisioner Pengujian Tesis Berjudul
"Analisis Jejaring Penelitian Kesehatan
Indonesia Menggunakan Text Mining"

Quisioner ni ditujukan untuk tenaga medis, mahasiswa, peneliti dan dosen di bidang kesehatan Mohon kesediaan nya untuk mengisi dengan jujur

* Wajib

Nama Lengkap *

Jawaban Anda

Insitusi *

Jawaban Anda

Apakah pemodelan pengetahuan Jejaring penelitian dengan menggunakan text mining yang dibangun mudah dipahami? *

Sangat Baik

Baik

Cukup

Kurang

Apakah pemodelan pengetahuan Jejaring penelitian dengan menggunakan text mining memudahkan para pengguna melakukan analisis penelitian kesehatan di Indonesia? *

- Sangat Baik
- Baik
- Cukup
- Kurang

Apakah pemodelan pengetahuan Jejaring penelitian dengan menggunakan text mining yang dihasilkan sudah sesuai? *

- Sangat Baik
- Baik
- Cukup
- Kurang

Apakah pemodelan pengetahuan Jejaring penelitian dengan menggunakan text mining bermanfaat bagi peneliti, tenaga kesehatan, dan akademis? *

- Sangat Baik
- Baik
- Cukup
- Kurang

Apakah sistem yang disediakan dapat mampu menunjang untuk mengembangkan penelitian kesehatan di Indonesia? *

- Sangat Baik
- Baik
- Cukup
- Kurang

Kirim