

**ANALISIS TREN TOPIK SISTEM INFORMASI DI INDONESIA DARI
PERSPEKTIF *TOPIC MODELING* MENGGUNAKAN LDA
(*LATENT DIRICHLET ALLOCATION*)**

SKRIPSI



**UIN SUNAN AMPEL
S U R A B A Y A**

Disusun Oleh:

REYNALDI OCTA PRADANA

H76216073

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL
SURABAYA**

2022

PERNYATAAN KEASLIAN

Saya yang bertanda tangan di bawah ini,

Nama : Reynaldi Octa Pradana

NIM : H76216073

Program Studi : Sistem Informasi

Angkatan : 2016

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul: “ANALISIS TREN TOPIK SISTEM INFORMASI DI INDONESIA DARI PERSPEKTIF *TOPIC MODELING* MENGGUNAKAN LDA (*LATENT DIRICHLET ALLOCATION*)”. Apabila suatu saat nanti terbukti saya melakukan tindakan plagiat, maka saya bersedia menerima sanksi yang telah ditetapkan. Demikian pernyataan keaslian ini saya buat dengan sebenar-benarnya.

Surabaya, 26 April 2023

Yang menyatakan,



(Reynaldi Octa Pradana)

NIM. H76216073

LEMBAR PERSETUJUAN PEMBIMBING

Skripsi oleh

NAMA : REYNALDI OCTA PRADANA
NIM : H76216073
JUDUL : ANALISIS TREN TOPIK SISTEM INFORMASI DI
INDONESIA DARI PERSPEKTIF *TOPIC MODELING*
MENGUNAKAN LDA (*LATENT DIRICHLET
ALLOCATION*)

Ini telah diperiksa dan disetujui untuk diujikan.

Surabaya, 7 April 2023

Dosen Pembimbing 1



Dwi Rolliawati, MT
NIP. 197909272014032001

Dosen Pembimbing 2



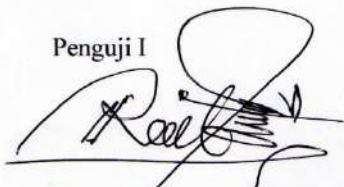
Khalid, M. Kom
NIP. 197906092014031002

PENGESAHAN TIM PENGUJI SKRIPSI

Skripsi Reynaldi Octa Pradana ini telah dipertahankan
di depan tim penguji skripsi
di Surabaya, 11 April 2023.

Mengesahkan,
Dewan Penguji

Penguji I



Mujib Ridwan, S.Kom., M.T
NIP. 198604272014031004

Penguji II



Ahmad Yusuf, M.Kom
NIP. 199001202014031003

Penguji III



Dwi Rolliawati, MT
NIP. 197909272014032001

Penguji IV




Khalid, M.Kom
NIP. 197906092014031002

Mengetahui,

Dekan Fakultas Sains dan Teknologi
Sunan Ampel Surabaya




Dr. Saepul Hamdani, M.Pd
NIP. 196507312000031002



UIN SUNAN AMPEL
SURABAYA

**KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA
PERPUSTAKAAN**

Jl. Jend. A. Yani 117 Surabaya 60237 Telp. 031-8431972 Fax.031-8413300
E-Mail: perpus@uinsby.ac.id

**LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademika UIN Sunan Ampel Surabaya, yang bertanda tangan di bawah ini, saya:

Nama : REYNALDI OCTA PRADANA
NIM : H76216073
Fakultas/Jurusan : SAINS DAN TEKNOLOGI / SISTEM INFORMASI
E-mail address : pradanareynaldi@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan UIN Sunan Ampel Surabaya, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah :

Skripsi Tesis Desertasi Lain-lain (.....)

yang berjudul :

ANALISIS TREN TOPIK SISTEM INFORMASI DI INDONESIA DARI PERSPEKTIF
TOPIC MODELING MENGGUNAKAN LDA (LATENT DIRICHLET ALLOCATION)

beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Eksklusif ini Perpustakaan UIN Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan, mengelolanya dalam bentuk pangkalan data (database), mendistribusikannya, dan menampilkan/mempublikasikannya di Internet atau media lain secara *fulltext* untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan UIN Sunan Ampel Surabaya, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini yang saya buat dengan sebenarnya.

Surabaya, 26 April 2023

Penulis

(REYNALDI OCTA PRADANA)

ABSTRAK

ANALISIS TREN TOPIK SISTEM INFORMASI DI INDONESIA DARI PERSPEKTIF *TOPIC MODELING* MENGGUNAKAN LDA (*LATENT DIRICHLET ALLOCATION*)

Oleh:

Reynaldi Octa Pradana

Peningkatan jumlah peneliti di Indonesia dari tahun ke tahun membuat jumlah publikasi penelitian ilmiah juga semakin meningkat. Selaras dengan hal tersebut, maka terdapat tren peningkatan publikasi ilmiah yang ada di Indonesia. Salah satu cara untuk menganalisis tren adalah menggunakan salah satu cabang keilmuan dari *data mining* yaitu metode *Topic Modeling*. Salah satu metode yang populer digunakan dalam *topic modeling* adalah metode LDA (*Latent Dirichlet Allocation*). Maka dari itu, penelitian ini meneliti tentang analisis tren topik sistem informasi di Indonesia. Dalam penelitian ini, data yang digunakan sebanyak 4354 baris dokumen dari total 30 jurnal yang dikumpulkan. Setelah itu data akan melalui proses *pre-processing*, kemudian pembentukan model oleh *Bag of Words*, dilanjutkan dengan mencari topik melalui metode LDA. Pada penelitian ini, metode *topic coherence* digunakan untuk mendapatkan jumlah topik pada model. Jumlah topik hasil dari *topic coherence* adalah 6 topik, namun karena saat divisualisasikan tidak membentuk model yang ideal, maka jumlah topik dikurangi hingga menjadi 3 topik. Pada penelitian ini, 3 topik tersebut direpresentasikan dengan topik Teknologi, topik Sistem Cerdas, dan topik Manajemen. Hasil dari penelitian ini adalah topik teknologi merupakan topik yang populer diangkat dalam penelitian sistem informasi, sedangkan topik Sistem Cerdas dan topik Manajemen masih jarang diangkat di dalam penelitian Sistem Informasi.

Kata Kunci: Sistem Informasi, Analisis Tren, *Topic Modeling*, LDA, *Topic Coherence*

ABSTRACT

TOPIC TREND ANALYSIS OF INFORMATION SYSTEMS IN INDONESIA FROM TOPIC MODELING PERSPECTIVE USING LDA (LATENT DIRICHLET ALLOCATION)

By:

Reynaldi Octa Pradana

The increase in the number of researchers in Indonesia yearly has led to an increase in the number of scientific researchers. In line with this, there is an increasing trend in scientific publications in Indonesia. One way to analyze trends is to use one of the scientific branches of data mining, which is the Topic Modeling method. One of the popular methods used in topic modeling is the LDA (Latent Dirichlet Allocation) method. Therefore, this research studied the trend analysis of information systems topics in Indonesia. In this research, the data used were 4354 lines of documents from a total of 30 journals collected. After that, the data will go through pre-processing, then model formation by Bag of Words, followed by finding topics through the LDA method. In this research, the topic coherence method is used to get the number of topics in the model. The total number of topics generated from topic coherence is 6 topics, but because the visualization does not form an ideal model, the number of topics is reduced to 3 topics. In this research, the 3 topics are represented by Technology topic, Intelligent System topic, and Management topic. The result of this research is that technology topics are popular topics to be discussed in information systems research, while Intelligent Systems topics and Management topics are still rarely discussed in Information Systems research.

Keyword: Information System, Trend Analysis, Topic Modeling, LDA, Topic Coherence

DAFTAR ISI

PERNYATAAN KEASLIAN.....	iii
LEMBAR PERSETUJUAN PEMBIMBING	iv
PENGESAHAN TIM PENGUJI SKRIPSI.....	v
KATA PENGANTAR	vi
ABSTRAK	viii
<i>ABSTRACT</i>	ix
DAFTAR ISI.....	x
DAFTAR TABEL.....	xii
DAFTAR GAMBAR	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah.....	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
1.6 Sistematika Penulisan.....	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 Tinjauan Penelitian Terdahulu	5
2.2 <i>Data Mining</i>	7
2.3 <i>Text Mining</i>	8
2.4 <i>Pre-Processing Data</i>	8
2.4.1 <i>Translating</i>	8
2.4.2 <i>Case Folding</i>	8
2.4.3 <i>Cleaning</i>	9
2.4.4 <i>Tokenizing</i>	9
2.4.5 <i>Lemmatization</i>	9
2.4.6 Penghapusan <i>Stopwords</i>	9
2.5 <i>Bag Of Words</i>	9
2.6 Pemodelan Topik.....	10
2.6.1 <i>Latent Dirichlet Allocation</i>	10
2.7 Integrasi Keilmuan	13

BAB III METODOLOGI PENELITIAN.....	15
3.1 Desain Penelitian	15
3.1.1 Pengumpulan Data.....	15
3.1.2 <i>Filtering & Remove Duplicates</i>	16
3.1.3 <i>Translating</i>	16
3.1.4 <i>Case Folding</i>	16
3.1.5 <i>Cleaning</i>	16
3.1.6 <i>Tokenizing</i>	17
3.1.7 <i>Lemmatization</i>	17
3.1.8 Penghapusan <i>Stopwords</i>	17
3.1.9 <i>Bag Of Words</i>	18
3.1.10 <i>Latent Dirichlet Allocation</i>	18
3.1.11 Pencarian Jumlah Topik	18
BAB IV HASIL & PEMBAHASAN.....	19
4.1 Hasil Dari Penelitian	19
4.1.1 Hasil Pengumpulan Data.....	19
4.1.2 Hasil Proses <i>Filtering</i> dan <i>Remove Duplicates</i>	24
4.1.3 Hasil <i>Pre-processing</i> Data.....	27
4.1.4 Hasil Proses Bag Of Words	36
4.1.5 Hasil Proses LDA.....	37
4.2 Pembahasan Analisis	48
4.2.1 Analisis Topik pada tahun 2015	48
4.2.2 Analisis Topik pada tahun 2016.....	50
4.2.3 Analisis Topik pada tahun 2017.....	51
4.2.4 Analisis Topik pada tahun 2018.....	52
4.2.5 Analisis Topik pada tahun 2019	53
4.2.6 Analisis Topik pada tahun 2020.....	54
4.2.7 Analisis Topik pada tahun 2021	55
4.3 Analisis Tren Tahunan	56
BAB V PENUTUP.....	59
5.1 Kesimpulan.....	59
5.2 Saran Pengembangan	59
DAFTAR PUSTAKA	60

DAFTAR TABEL

Tabel 2. 1 Daftar Tinjauan Pustaka	5
Tabel 3. 1 Contoh dari proses Case Folding	16
Tabel 3. 2 Contoh dari proses Cleaning	17
Tabel 3. 3 Contoh dari proses Tokenizing	17
Tabel 3. 4 Contoh dari proses Lemmatization	17
Tabel 3. 5 Contoh dari proses Penghapusan Stopwords	18
Tabel 4. 1 Indeks Jurnal dan Jurnal yang digunakan sebagai data penelitian	19
Tabel 4. 2 Hasil dari proses Filtering dan Remove Duplicates	24
Tabel 4. 3 Hasil dari Proses Translating.....	28
Tabel 4. 4 Hasil dari Proses Case Folding.....	29
Tabel 4. 5 Hasil dari Proses Cleaning	30
Tabel 4. 6 Hasil dari proses Tokenizing.....	32
Tabel 4. 7 Hasil dari prsoes Lemmatization.....	33
Tabel 4. 8 Hasil dari proses penghapusan Stopwords	35
Tabel 4. 9 Hasil dari proses Bag of Words.....	36
Tabel 4. 10 Hasil Proses LDA Data tahun 2015	41
Tabel 4. 11 Hasil Proses LDA Data tahun 2016	42
Tabel 4. 12 Hasil Proses LDA Data tahun 2017	43
Tabel 4. 13 Hasil Proses LDA Data tahun 2018	44
Tabel 4. 14 Hasil Proses LDA Data tahun 2019	45
Tabel 4. 15 Hasil Proses LDA Data tahun 2020	46
Tabel 4. 16 Hasil Proses LDA Data tahun 2021	47
Tabel 4. 17 Hasil interpretasi topik tahun 2015	49
Tabel 4. 18 Hasil interpretasi topik tahun 2016	50
Tabel 4. 19 Hasil interpretasi topik tahun 2017	51
Tabel 4. 20 Hasil interpretasi topik tahun 2018	52
Tabel 4. 21 Hasil interpretasi topik tahun 2019	53
Tabel 4. 22 Hasil interpretasi topik tahun 2020	54
Tabel 4. 23 Hasil interpretasi topik tahun 2021	55
Tabel 4. 24 Jumlah dokumen terkait ke tiap topik di dalam corpus.....	56

DAFTAR GAMBAR

Gambar 2. 1 Langkah-Langkah dalam Data Mining	7
Gambar 2. 2 Konsep LDA oleh Blei	11
Gambar 2. 3 Representasi Model LDA	11
Gambar 2. 4 Langkah-langkah proses Topic Coherence.....	12
Gambar 3. 1 Desain Penelitian	15
Gambar 4. 1 Selector Graph untuk proses Scrapping Abstrak.....	19
Gambar 4. 2 Persentase Data Hasil Scrapping petahun.....	23
Gambar 4. 3 Persentase data hasil proses filtering dan remove duplicates.....	27
Gambar 4. 4 Hasil Topic Coherence.....	38
Gambar 4. 5 Pemetaan topik LDA dengan jumlah topik 6.....	39
Gambar 4. 6 Pemetaan topik LDA dengan jumlah topik 5.....	39
Gambar 4. 7 Pemetaan topik LDA dengan jumlah topik 4.....	40
Gambar 4. 8 Pemetaan topik LDA dengan jumlah topik 3.....	40
Gambar 4. 9 Hasil visualisasi proses LDA corpus 2015.....	42
Gambar 4. 10 Hasil visualisasi proses LDA corpus 2016.....	43
Gambar 4. 11 Hasil visualisasi proses LDA corpus 2017.....	44
Gambar 4. 12 Hasil visualisasi proses LDA corpus 2018.....	45
Gambar 4. 13 Hasil visualisasi proses LDA corpus 2019.....	46
Gambar 4. 14 Hasil visualisasi proses LDA corpus 2020.....	47
Gambar 4. 15 Hasil visualisasi proses LDA corpus 2021.....	48
Gambar 4. 16 Persentase jumlah dokumen terhadap topik di 2015.....	49
Gambar 4. 17 Persentase jumlah dokumen terhadap topik di 2016.....	50
Gambar 4. 18 Persentase jumlah dokumen terhadap topik di 2017.....	51
Gambar 4. 19 Persentase jumlah dokumen terhadap topik di 2018.....	52
Gambar 4. 20 Persentase jumlah dokumen terhadap topik di 2019.....	53
Gambar 4. 21 Persentase jumlah dokumen terhadap topik di 2020.....	54
Gambar 4. 22 Persentase jumlah dokumen terhadap topik di 2021.....	55
Gambar 4. 23 Tren tahunan topik Sistem Cerdas.....	56
Gambar 4. 24 Tren tahunan topik Teknologi.....	57
Gambar 4. 25 Tren tahunan topik Manajemen.....	57

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kenaikan angka jumlah peneliti di Indonesia setiap tahunnya membuat angka publikasi penelitian ilmiah juga semakin meningkat. Pada tahun 2015 jumlah peneliti di Indonesia dari 9.308 naik menjadi 9.685 di tahun 2017 (Gerintya, 2017). Selaras dengan perkembangan bidang publikasi penelitian ilmiah di Indonesia, menganalisis tren penelitian merupakan hal yang umum dilakukan. Hal tersebut dilakukan untuk mengetahui bagaimana gambaran *tren* penelitian dari waktu ke waktu (Nastiti, 2019). Salah satu metode yang bisa digunakan untuk menganalisa *tren* penelitian yaitu dengan menggunakan *Data Mining*.

Data Mining adalah sebuah teknologi yang banyak dimanfaatkan oleh perusahaan-perusahaan untuk mendapatkan informasi yang penting dari *data warehouse* perusahaan tersebut (Moertini, 2002). Penggunaan *Data Mining* tidak terbatas untuk perusahaan saja, salah satu penggunaan *Data Mining* adalah Pemodelan Topik. Menurut penelitian (Putra, 2017) Pemodelan Topik merupakan algoritma untuk mendapatkan topik-topik yang tidak terlihat melalui susunan kata di dalam dokumen yang tidak berstruktur. Pemodelan Topik berfokus untuk mendapatkan topik dalam kumpulan data dokumen.

Di dalam Pemodelan Topik, terdapat banyak metode yang bisa digunakan contohnya *Probabilistic Latent Semantic Analysis* (PLSA), *Latent Semantic Analysis* (LSA), dan *Latent Dirichlet Allocation* (LDA). LDA adalah algoritma untuk menemukan topik dengan pemodelan probabilistik di dalam kumpulan data. LDA merupakan metode Pemodelan Topik yang paling populer karena dapat melakukan analisis pada dokumen yang ukurannya besar. LDA menghasilkan kumpulan topik yang memiliki tiap topiknya memiliki bobot masing-masing di dalam dokumen sehingga LDA biasanya digunakan untuk melakukan *clustering*, serta menghubungkan ataupun memproses data yang ukurannya sangat besar (Putra, 2017).

Anand Jerayaj dan Amir Hassan Zadeh melakukan penelitian terhadap perkembangan riset Sistem Informasi selaras dengan berkembangnya keilmuan Sistem Informasi. Data yang digunakan merupakan data penelitian ilmiah yang terpublikasi secara *online* di jurnal Sistem Informasi antara tahun 2013-2017 sebanyak 2962 artikel penelitian. Jurnal Sistem Informasi yang digunakan sebagai data penelitian adalah *Information System Research (ISR)*, *European Journal of Information System (EJIS)*, *Journal of Management Information System (JMIS)*, *Information System Journal (IJS)*, *Journal of the Association for Information Systems (JAIS)*, dan *MIS Quarterly (MISQ)*. Penelitian tersebut menghasilkan tren dari 35 topik penelitian SI yang sudah dimodelkan (Jeyaraj & Zadeh, 2019).

Penelitian dengan menggunakan metode LDA ini sudah banyak dilakukan oleh beberapa peneliti di Indonesia dengan menggunakan studi kasus lain. Penelitian yang dilakukan oleh (Zulhanif, 2016) mengambil data sampel dari *platform social media* twitter dengan penyaringan yang mengandung kata #Bandung, dan dari penyaringan tersebut didapatkan data sebanyak 1500 *tweet*. Metode pengklasteran dengan metode LDA pada penelitian ini menghasilkan 24 topik yang diimplementasikan dalam Bahasa R. Keterbatasan pada penelitian ini adalah jumlah dimensi dari *term* yang terbentuk. Penelitian lain dilakukan oleh (Utami, 2017) meneliti tentang *tweet* pada 5 daerah di kota Bogor yaitu Bogor Selatan, Bogor Barat, Bogor Utara, Bogor Kota, dan Bogor Timur selama 1 minggu. Jumlah *tweet* yang digunakan sebanyak 7.387 *tweet*. Penelitian menghasilkan pembentukan topik dari informasi di setiap *tweetnya*. Informasi tersebut dapat oleh *stakeholder* untuk memahami perkembangan isu terkini. Oleh karena hal tersebut, dalam penelitian ini metode LDA digunakan karena metode LDA adalah salah satu metode dalam pemodelan topik yang populer saat ini. Selain itu, LDA juga memiliki kelebihan untuk memproses data yang berukuran sangat besar (Alfanzar, 2019).

Oleh karena hal itu, pada penelitian ini akan mengulas tentang analisis tren topik system informasi di Indonesia yang *publish* secara *online* di internet pada kurun waktu tahun 2015 hingga tahun 2021 dengan memanfaatkan metode Pemodelan Topik LDA.

1.2 Perumusan Masalah

Berdasar kepada latar belakang yang sudah disusun sebelumnya, maka untuk perumusan masalah dalam penelitian ini adalah bagaimana analisis tren penelitian Sistem Informasi di Indonesia dalam kurun waktu 2015 hingga 2021 dengan menggunakan Pemodelan Topik metode LDA?

1.3 Batasan Masalah

Batasan masalah pada penelitian ini yaitu:

1. Data penelitian yang diambil adalah artikel-artikel pada jurnal yang terindeks oleh Sinta dengan nilai indeks S1 – S4 pada situs Sinta.kemdikbud.go.id .
2. Data yang diambil merupakan data dari artikel-artikel yang jurnalnya tersedia di situs Garuda.kemdikbud.go.id
3. Data yang diambil merupakan isi abstrak artikel dalam jurnal penelitian yang terpublikasi secara *online* pada jurnal dengan kata kunci pencarian judul jurnal “Sistem Informasi” dan “*Information System*”.
4. Data yang diambil merupakan abstrak dari artikel yang *terpublish* dalam kurun waktu 2015 sampai 2021.
5. Bahasa *programming* yang digunakan pada penelitian ini adalah Bahasa *Python*.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini yaitu untuk mendapatkan tren riset Sistem Informasi yang ada di Indonesia pada kurun waktu tahun 2015 hingga tahun 2021 dengan mengimplementasikan metode Pemodelan Topik LDA.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah:

1. Manfaat di sisi Akademis: Melihat tren penelitian Sistem Informasi di Indonesia selama 6 tahun terakhir.
2. Manfaat di sisi Praktis: Sebagai bahan rujukan untuk peneliti di Indonesia dalam menentukan topik penelitian pada keilmuan Sistem Informasi di masa depan.

1.6 Sistematika Penulisan

Sistematika penulisan skripsi yang telah disusun oleh Program Studi Sistem Informasi Fakultas Sains dan Teknologi UIN Sunan Ampel Surabaya yaitu sebagai berikut:

a. BAB I PENDAHULUAN

Pada Bab I menjelaskan tentang latar belakang masalah yang mendasari analisis tren riset sistem informasi ini dilakukan serta metode yang akan digunakan dalam penelitian ini. Selain itu, pada Bab I juga dijelaskan rumusan masalah, batasan masalah, tujuan penelitian, manfaat analisis tren riset sistem informasi.

b. BAB II TINJAUAN PUSTAKA

Pada Bab II menjelaskan tentang penelitian terdahulu yang memiliki hubungan dengan penelitian yang diteliti seperti penelitian-penelitian yang menggunakan metode yang sama (LDA) dan penelitian yang mengambil objek penelitian yang sama (tren riset). Selain itu juga pada Bab II dijelaskan dasar teori dari pengetahuan-pengetahuan yang digunakan di dalam penelitian seperti Data Mining, Pemodelan Topik, LDA, dll, serta integrasi keilmuan antara penelitian dengan Islam.

c. BAB III METODOLOGI PENELITIAN

Pada Bab III menjelaskan tentang alur proses dalam penelitian dimulai dari pengumpulan data hingga dilakukan analisis hasil penelitian. Pada Bab III juga dijelaskan tentang *timeline* penelitian dilakukan dari proses awal hingga akhir.

d. BAB IV HASIL DAN PEMBAHASAN

Pada Bab IV menjelaskan tentang hasil dari setiap alur proses penelitian dimulai dari pengambilan data, kemudian *pre-processing* data, setelah itu proses LDA untuk mendapatkan bahan analisis riset Sistem Informasi. Kemudian dilanjutkan dengan proses mencari nilai *topic popularity* untuk mendapat tren analisis riset.

e. BAB V PENUTUP

Pada Bab V menjelaskan tentang kesimpulan dari penelitian dan juga saran yang ditunjukkan kepada pembaca atau peneliti selanjutnya, sehingga penelitian ini dapat dikembangkan lagi di masa mendatang.

BAB II

TINJAUAN PUSTAKA

2.1 Tinjauan Penelitian Terdahulu

Tinjauan penelitian terdahulu bertujuan untuk memperkuat pemahaman tentang penelitian ini. Penelitian yang diambil sebagai tinjauan penelitian terdahulu adalah penelitian yang memiliki metode (LDA) atau objek penelitian (tren riset) yang sama seperti yang terdapat pada Tabel 2.1 berikut.

Tabel 2. 1 Daftar Tinjauan Pustaka

No.	Judul Penelitian	Hasil	Korelasi
1.	Evolution of Information System Research : Insights from Topic Modeling (Jeyaraj & Zadeh, 2019)	Penelitian dilakukan untuk memetakan perkembangan penelitian Sistem Informasi. Penelitian menggunakan artikel yang <i>terpublish</i> pada 6 jurnal Sistem Informasi internasional teratas pada tahun 2013-2017. <i>Topic modeling</i> dilakukan kepada 600 lebih kata kunci unik yang ada di artikel penelitian. Penelitian menghasilkan perkembangan penelitian Sistem Informasi yang <i>upward trend</i> (naik) maupun yang <i>downward trend</i> (turun).	Penelitian ini meneliti tentang tren perkembangan penelitian SI di Eropa menggunakan metode LSA.
2.	<i>Topic Modelling</i> Skripsi Menggunakan Metode <i>Latent Dirichlet Allocation</i> (Alfanzar, 2019)	Proses implementasi <i>topic modelling</i> di <i>abstract</i> skripsi program studi Sastra Inggris di UINSA dilakukan dengan percobaan sebanyak 4 uji iterasi yang berbeda (100, 500, 1000, 5000) dan 5 topik yang berbeda (2,3,4,5,7 topik). Hasil <i>cluster</i> terbaik yaitu terdapat pada topik 3. Hasil telah diverifikasi dan	Penelitian menggunakan metode LDA namun tidak terdapat proses validasi model

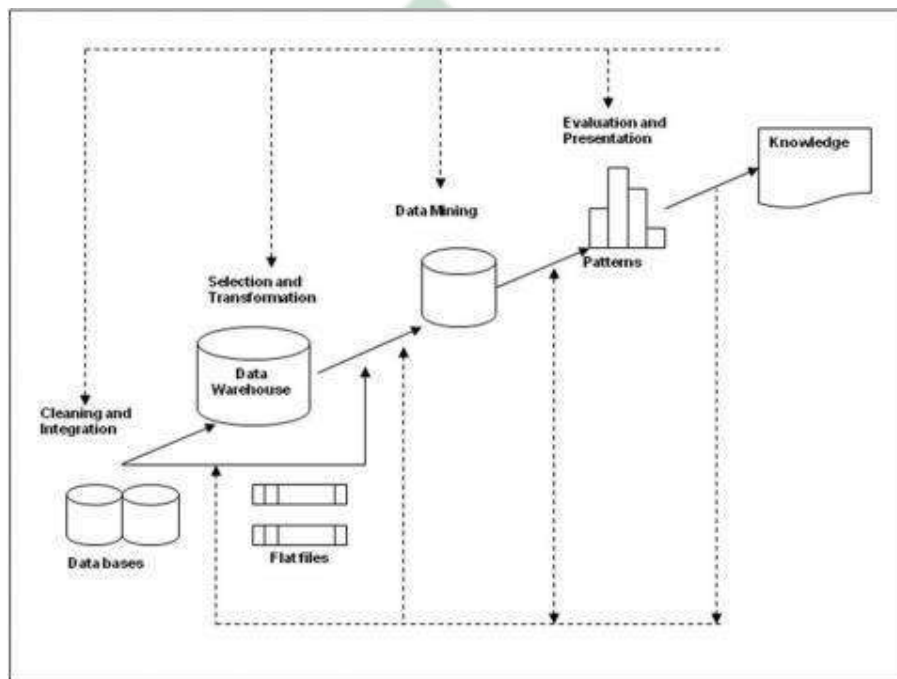
		dinyatakan sesuai oleh pihak terkait di program studi Sastra Inggris di UINSA.	
3.	Pemodelan Topik Dengan <i>Latent Dirichlet Allocation</i> (Zulhanif, 2016)	Metode <i>clustering</i> LDA pada data <i>twitter</i> dengan menggunakan kata kunci #Bandung menghasilkan 24 buah topik.	Penelitian LDA ini menggunakan data penelitian yang berasal dari twitter.
4.	Analisis topik pada suara konsumen PT Angkasa Pura 1 (Persero) cabang Juanda dengan menggunakan pemodelan Gaussian Latent Dirichlet Allocation (GLDA) (Rozaliya, 2018)	Hasil dari eksperimen, didapatkan model LDA terbaik adalah model dengan jumlah topik 4 dengan skenario data Tanpa Frasa dan <i>stemming</i> batasan 1 serta memiliki nilai <i>topic coherence</i> -5.326.	Penelitian ini meneliti data dari suara konsumen dengan menggunakan metode LDA
5.	Analisis Topik Data Media Sosial Twitter Menggunakan Model Topik Latent Dirichlet Allocation (Utami, 2017)	Hasil dari pemodelan topik LDA pada 5 lokasi <i>tweet</i> di Kota Bogor dan rentang waktu tertentu berhasil membentuk topik dengan informasi untuk setiap topik di setiap lokasi <i>tweet</i> . Informasi dalam topik dapat digunakan oleh pembaca ataupun <i>stakeholder</i> yang terkait untuk memahami setiap perkembangan isu terkini.	Penelitian ini meneliti data media sosial dengan metode LDA.

Berdasarkan tinjauan dari penelitian terdahulu di Tabel 2.1, penelitian *text mining* yang memanfaatkan metode Pemodelan Topik *Latent Dirichlet Allocation* (LDA) cukup banyak. Penelitian banyak dilakukan kepada data-data media sosial dan juga berita. Data yang digunakan dalam penelitian ini adalah data dari artikel-artikel penelitian yang ada di dalam jurnal Sistem Informasi di Indonesia. Dalam penelitian ini juga digunakan metode *topic coherence* untuk memvalidasi model

yang akan digunakan. Analisis lanjutan dilakukan untuk mendapatkan tren pada topik sistem informasi di Indonesia.

2.2 Data Mining

Data Mining merupakan proses mengekstrak *knowledge* yang berasal dari data dengan ukuran besar untuk menemukan informasi penting yang dapat berguna untuk pengembangan. Pada Gambar 2.1 menjelaskan tentang langkah yang dilakukan pada *data mining*.



Gambar 2. 1 Langkah-Langkah dalam *Data Mining*

Sumber: (Fadli, 2003)

Berikut merupakan penjelasan dari Gambar 2.1 di atas:

1. *Data Cleaning* (proses membersihkan data inkonsisten)
2. *Data Integration* (proses menyatukan sumber data yang terpisah)
3. *Data Selection* (proses menganalisis data yang relevan dengan tugas)
4. *Data Transformation* (proses pengubahan data ke bentuk yang tepat agar dapat diproses)
5. *Data Mining* (proses mengolah pola data yang ada)
6. *Pattern Evolution* (proses pengidentifikasian pola untuk mendapatkan pengetahuan)
7. *Knowledge presentation* (proses menggambarkan pengetahuan yang telah diproses dengan teknik visualisasi kepada user).

2.3 Text Mining

Text mining yaitu salah satu cabang dari keilmuan *data mining*. Metode *Text mining* adalah proses menggali *text* yang digunakan sebagai data dan datanya bersumber dari dokumen (pdf, word, dan sejenisnya). *Text mining* dapat disebut juga dengan teknik untuk mengekstraksi pola data dari sebuah dokumen. Tujuan utama dari metode *text mining* yaitu untuk menemukan kumpulan kata dalam sekumpulan dokumen yang merepresentasikan isi di dalam dokumen tersebut dan diolah menjadi sebuah informasi yang berguna untuk dianalisa lebih lanjut. *Text mining* dapat digunakan sebagai pengklasifikasian teks (*text classification*) dan pengelompokan teks (*text clustering*) (Aliyanto, 2017).

2.4 Pre-Processing Data

Tahapan pra-proses data atau *preprocessing* data adalah tahapan untuk mempersiapkan data *raw* (mentah) yang akan digunakan sebelum melakukan tahapan lainnya (Mujilahwati, 2016). Penelitian lain menyebutkan bahwa *preprocessing* adalah tahap proses awal pada *text mining* yang dilakukan untuk tujuan mempersiapkan data agar dapat diolah sesuai dengan kebutuhan (Aditya, 2015).

2.4.1 Translating

Translating merupakan suatu pengganti bahan tekstual dari suatu bahasa yang adalah Bahasa sumber atau *source language* oleh bahan tekstual yang setara dalam bahasa sasaran atau *target language* (Catford, 1965). Proses *translating* pada *pre-processing* bertujuan untuk menyamakan data-data yang berbeda bahasa menjadi satu bahasa yang sama tanpa membuat perbedaan dalam arti atau makna. Proses *translating* dilakukan untuk memudahkan proses yang dilakukan setelahnya.

2.4.2 Case Folding

Case Folding yaitu proses untuk menyeragamkan bentuk huruf pada suatu data. Data-data yang ada bentuknya diubah menjadi huruf kecil. Dalam proses ini hanya menerima bentuk data yang berupa huruf latin antara a-z. (Putranti & Winarko, 2014).

2.4.3 Cleaning

Cleaning merupakan proses mendeteksi dan juga menghapuskan *error* serta inkonsistensi yang ada di dalam sebuah data. Proses *cleaning* bertujuan untuk meningkatkan kualitas data sehingga bisa digunakan secara maksimal untuk proses selanjutnya (Rahm & Do, 2000).

2.4.4 Tokenizing

Menurut (Triawati, 2009) *tokenizing* adalah proses pengidentifikasian unit terkecil dari sebuah struktur kalimat. Pada penelitian (Sasmita, 2014), *tokenizing* didefinisikan sebagai proses pemisahan *string* berdasarkan kata-kata yang menyusunnya. Jadi sebuah kalimat dirubah menjadi ke bentuk kata-kata yang menyusunnya.

2.4.5 Lemmatization

Lemmatization adalah proses untuk menguraikan berbagai bentuk kata menjadi ke bentuk dasarnya. Proses penguraian ini dilakukan untuk mendapatkan bentuk dasar dari tiap kata yang berimbuhan. Proses ini dilakukan dengan cara menghapus imbuhan yang terdapat pada kata tersebut (Afuan, 2013).

2.4.6 Penghapusan Stopwords

Stopwords adalah sekumpulan kata di dalam suatu dokumen *corpus* yang memiliki frekuensi kemunculan berbeda dengan kebanyakan kata yang lainnya. Kumpulan kata tersebut tidak mempunyai makna yang berarti pada dokumen. Contoh kata *stopwords* dalam Bahasa Inggris adalah “*to*”, “*from*”, “*this*”, dst. Kata-kata tersebut merupakan kata-kata penghubung di dalam sebuah kalimat yang sering muncul pada dokumen-dokumen berbahasa Indonesia (Utomo, 2015).

2.5 Bag Of Words

Bag Of Words (BoW) adalah salah satu model yang merepresentasikan data secara sederhana dan populer digunakan untuk metode *Natural Language Processing* (NLP) dan juga metode *Information Retrieval* (IR). Dalam model BoW ini, *text* yang bentuknya kalimat atau dokumen direpresentasikan sebagai *bag multiset* dari kata-kata yang ada di dalamnya, tanpa memperhatikan tata bahasa serta urutan kata tetapi tetap mempertahankan keberagamannya (Putri & Hendrowati, 2018). Definisi lain untuk model BoW adalah sebuah model yang

mempelajari tiap kosakata yang terdapat di semua dokumen, kemudian menghitung frekuensi kemunculan setiap kata di dalamnya untuk mendapatkan model dari tiap dokumennya. (Deepu et al., 2016).

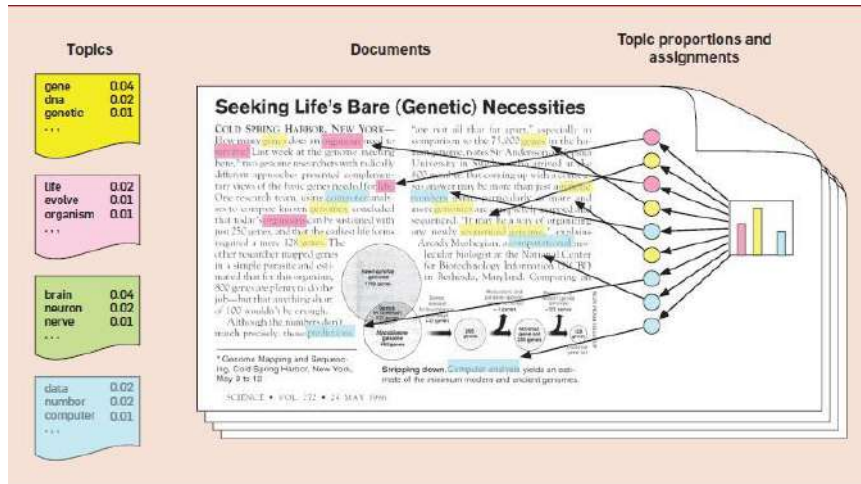
2.6 Pemodelan Topik

Konsep pemodelan topik menurut (Blei et al., 2003) adalah sebuah konsep yang di dalamnya terdapat banyak entitas seperti “*word*”, “*document*”, dan “*corpora*”. “*Word*” dianggap sebagai unit terkecil pada dokumen. Sedangkan “*document*” adalah susunan dari tiap kata yang ada. *Corpus* merupakan sekumpulan dokumen, sedangkan *corpora* adalah bentuk jamaknya *corpus* itu sendiri. “*Topic*” merupakan distribusi kosakata yang sifatnya tetap. Jadi, terdapat banyak topik yang memiliki proporsi tersendiri sesuai dengan kata-kata yang menyusunnya di dalam tiap dokumen di *corpus*.

Secara pemahaman manusia, topik serta distribusi topik di setiap dokumen, dan juga penggolongan kata ke dalam topik di tiap dokumennya merupakan struktur yang tidak bisa diamati atau tersembunyi. Oleh karena hal tersebut, pemodelan topik dilakukan untuk mendapatkan hal-hal yang tidak bisa diamati oleh manusia di dalam dokumen-dokumen tersebut.(Putra, 2017).

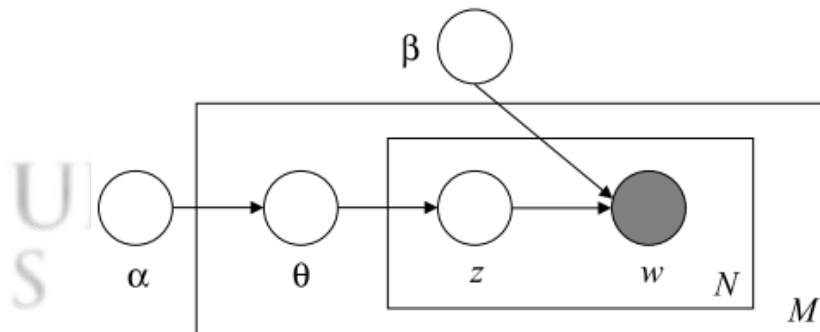
2.6.1 Latent Dirichlet Allocation

Latent Dirichlet Allocation yaitu salah satu dari banyaknya metode dalam pemodelan topik yang populer digunakan saat ini. Menurut (Blei et al., 2003) LDA adalah model probabilistik generatif dari sekumpulan *corpus*. LDA ini tidak terikat hanya dengan data teks, namun juga dapat diaplikasikan kepada permasalahan lain seperti dalam mengumpulkan data, *bioinformatics*, dll. Pada metode LDA, suatu dokumen dikenali sebagai model yang bercampur dari banyak macam topik yang bisa disebut juga sebagai *latent*, yang mana tiap topik tersebut dikarakteristikan oleh kumpulan kata yang menyusunnya(Zulhanif, 2016). Konsep Pemodelan topik LDA oleh David M. Blei Digambarkan seperti Gambar 2.2.



Gambar 2. 2 Konsep LDA oleh Blei
 Sumber: (Putra, 2017)

Dalam setiap dokumen terdapat kumpulan kata yang secara acak dapat membentuk suatu topik. Kumpulan kata tersebut memiliki nilai proporsinya masing-masing di dalam dokumen tersebut sehingga kata-kata tersebut membentuk suatu topik. Di dalam suatu dokumen, kumpulan kata tersebut dapat membentuk lebih dari 1 topik sehingga dapat disimpulkan bahwa di dalam dokumen tidak hanya terdapat 1 topik melainkan banyak topik. Model LDA jika direpresentasikan dalam bentuk gambar diagram seperti Gambar 2.3 berikut:



Gambar 2. 3 Representasi Model LDA
 Sumber: (Blei et al., 2003)

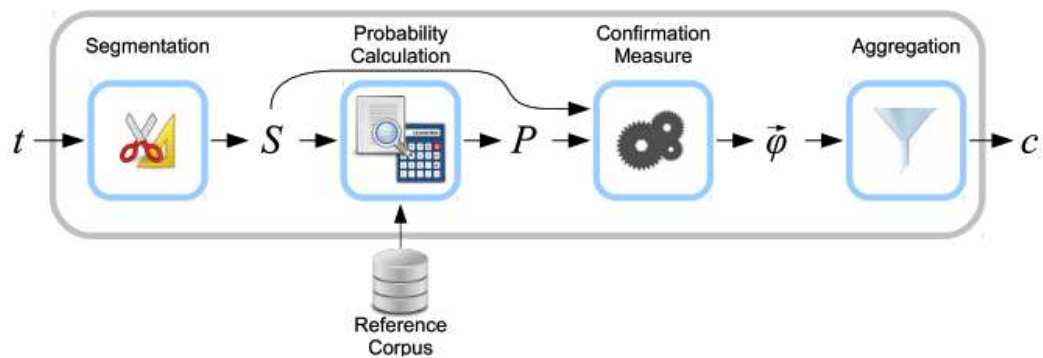
Keterangan diagram sebagai berikut:

- α = Parameter untuk distribusi *topic* yang berada dalam tahapan *corpus*. Berfungsi untuk memutuskan distribusi topik pada sebuah dokumen. Apabila nilai α pada suatu dokumen semakin besar maka menandakan semakin banyak campuran topik di dalam dokumen tersebut.

- β = Parameter untuk mendapatkan nilai distribusi kata pada suatu topik. Semakin besar nilai dari β artinya makin banyak kata yang menyusun suatu topik.
- θ = Variabel yang berada pada tingkat dokumen. Semakin tinggi nilai θ menandakan dokumen-dokumen tersebut memiliki topik yang banyak.
- z_N = Variabel ini merepresentasikan suatu topik dari sebuah kata di dalam dokumen.
- w_N = Variabel yang digunakan untuk merepresentasikan kata yang terkait kepada suatu topik tertentu di dalam dokumen.

2.6.2 Topic Coherence

Evaluasi dengan menggunakan nilai *topic coherence* bertujuan untuk menilai seberapa mudah kumpulan topik yang didapatkan dari model untuk dipahami oleh manusia. Pengukuran ini berfungsi untuk membantu membedakan antara topik yang berkaitan secara statistic dengan topik yang dapat dikenali secara semantik (Putra, 2017). Langkah-langkah yang terjadi dalam *topic coherence* seperti di Gambar 2.4 berikut:



Gambar 2. 4 Langkah-langkah proses *Topic Coherence*
 Sumber: (Röder et al., 2015)

Langkah pertama adalah proses segmentasi pada kumpulan topik t sehingga menjadi kumpulan *set* pasangan dari *subset* kata S . Setelah itu melakukan penghitungan probabilitas untuk tiap kata P berdasarkan data kuantitatif dari *corpus*. Selanjutnya, nilai P dan S diolah pada proses *confirmation measure* yang bertujuan untuk menghitung nilai syarat kualitas dari pasangan S menghasilkan ϕ . Terakhir, semua nilai $\phi = \{ \phi_1, \dots, \phi/S \}$ dihitung rata-ratanya sehingga menghasilkan nilai c yaitu nilai *coherence* (Röder et al., 2015).

2.7 Integrasi Keilmuan

Pemodelan topik adalah salah satu bagian dari keilmuan Data Mining yang bermanfaat untuk mencari informasi dari kumpulan dokumen atau data yang tidak terstruktur. Informasi tersebut kemudian dapat diolah menjadi sebuah pengetahuan yang penting dan dapat dimanfaatkan dengan baik.

Berdasarkan wawancara dengan Ustadz Didi Sunardi, beliau mengatakan bahwa di dalam Islam juga terdapat ayat yang menganjurkan manusia untuk mencari informasi yaitu pada Q.S. Al-Alaq ayat 1-5 dan Q.S. Al-Hujurat ayat 6 yang di dalamnya berisi tentang pentingnya untuk menggali ilmu pengetahuan.

إِقْرَأْ بِاسْمِ رَبِّكَ الَّذِي خَلَقَ ١ - خَلَقَ الْإِنْسَانَ مِنْ عَلَقٍ ٢ - اقْرَأْ وَرَبُّكَ الْأَكْرَمُ ٣ - الَّذِي عَلَّمَ بِالْقَلَمِ ٤ - عَلَّمَ الْإِنْسَانَ مَا لَمْ يَعْلَمْ ٥

Yang artinya:

(1) Bacalah dengan (menyebut) nama Tuhanmu yang menciptakan, (2) Dia telah menciptakan manusia dari segumpal darah, (3) Bacalah, dan Tuhanmulah yang Maha Pemurah, (4) Yang mengajar (manusia) dengan perantaran kalam, (5) Dia mengajar kepada manusia apa yang tidak diketahuinya.

يَا أَيُّهَا الَّذِينَ آمَنُوا إِن جَاءَكُمْ فَاسِقٌ بِنَبَأٍ فَتَبَيَّنُوا أَن تُصِيبُوا قَوْمًا بِجَهَالَةٍ فَتُصْحَبُوا عَلَىٰ مَا فَعَلْتُمْ نَادِمِينَ ٦

Yang artinya:

(6) Wahai orang-orang yang beriman! Jika seseorang yang fasik datang kepadamu membawa suatu berita, maka telitilah kebenarannya, agar kamu tidak mencelakakan suatu kaum karena kebodohan (kecerobohan), yang akhirnya kamu menyesali perbuatanmu itu.

Pada Q.S. Al-Alaq di atas, diceritakan bahwa Allah menghimbau umat manusia untuk membaca, karena dengan membaca maka manusia akan mendapatkan informasi ataupun pengetahuan. Sedangkan pada Q.S. Al-Hujurat di

ayat 6, Allah SWT menghimbau manusia untuk menggali kebenaran atau informasi yang diterimanya agar tidak mencelakai orang lain dan menyesal. Semakna dengan tujuan dari Pemodelan Topik yang mana tujuannya untuk mencari sebuah topik informasi dari kumpulan dokumen yang tidak teratur. Pemodelan topik dilakukan untuk mendapatkan sebuah informasi yang ada pada sebuah dokumen dalam berbentuk topik.



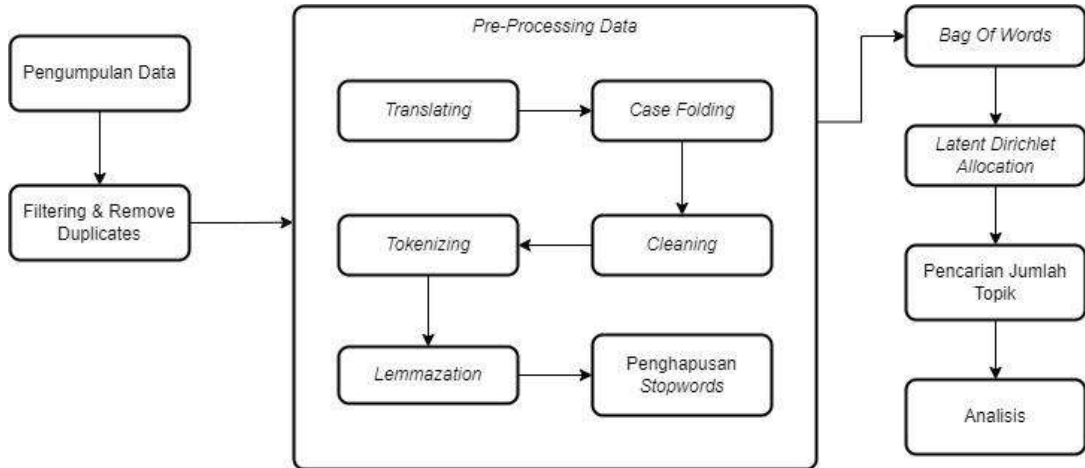
UIN SUNAN AMPEL
S U R A B A Y A

BAB III

METODOLOGI PENELITIAN

3.1 Desain Penelitian

Desain dari penelitian ini dijelaskan dalam bentuk *flowchart* untuk mempermudah dalam memahami alur penelitian. Gambar 3.1 menampilkan desain dari penelitian ini.



Gambar 3.1 Desain Penelitian

Berikut merupakan penjelasan dari setiap proses di dalam desain penelitian.

3.1.1 Pengumpulan Data

Penelitian ini menggunakan data yang diambil dari situs Garuda.kemdikbud.go.id yaitu website penyedia informasi tentang jurnal-jurnal berbahasa Indonesia yang *terpublish* secara *online*. Adapun data yang diambil merupakan data dari artikel penelitian tahun 2015 hingga 2021 yang merupakan bagian dari jurnal dengan *keyword* pencarian Sistem Informasi / *Information System* yang terindeks oleh Sinta dengan nilai indeks Sinta 1, Sinta 2, Sinta 3, hingga Sinta 4. Informasi yang diambil adalah tahun penelitian dan abstrak penelitian. Pengumpulan data untuk penelitian ini dilakukan dengan mengandalkan bantuan *tool* dari *Google Chrome* yaitu *Web Scrapper extension*.

Dalam melakukan proses pengumpulan data menggunakan *tool Web Scrapper extension*, data yang diambil melalui beberapa proses seperti berikut:

- a. Mengakses jurnal di *website* yang dituju;
- b. Membuat *sitemap* dari jurnal yang ingin diambil informasinya;

- c. Membuat *selector* pada informasi-informasi yang ingin *discrapping*;
- d. Menghubungkan *selector* dengan *parent selector* agar informasi yang diambil merupakan satu kesatuan;
- e. Melakukan proses *scrapping* pada sitemap jurnal;
- f. *Export* hasil proses *scrapping* menjadi dokumen *xlsx*..

3.1.2 Filtering & Remove Duplicates

Proses pertama sebelum melakukan *pre-processing* adalah melakukan *filtering* untuk data-data yang tidak termasuk ke dalam batasan penelitian. Selain itu juga dilakukan penghilangan duplikat agar data yang akan digunakan nantinya dapat menghasilkan hasil yang maksimal saat diproses.

3.1.3 Translating

Proses selanjutnya pada tahap *pre-processing* adalah menerjemahkan data yang digunakan menjadi 1 bahasa yang sama. Pada tahap *translating*, data-data penelitian yang tertulis dengan bahasa yang berbeda akan diubah menjadi satu bahasa yang sama agar memudahkan proses selanjutnya. Pada tahap *translating* ini data penelitian yang menggunakan bahasa Indonesia akan diubah menjadi bahasa Inggris tanpa mengubah makna dari kata tersebut.

3.1.4 Case Folding

Proses selanjutnya yaitu membuat semua data yang telah dikumpulkan menjadi menjadi seragam. Proses yang termasuk di dalam *case folding* tersebut bertujuan supaya kata yang sama tetapi penulisannya berbeda tidak dianggap sebagai dua kata berbeda. Contoh dari proses *case folding* terdapat di Tabel 3.1.

Tabel 3. 1 Contoh dari proses *Case Folding*

Sebelum	Sesudah
Saya pernah pergi ke banyak kota diantaranya kota Bogor, kota Semarang, dan kota Surabaya	saya pernah pergi ke banyak kota diantaranya kota bogor, kota semarang, dan kota surabaya

3.1.5 Cleaning

Proses pertama pada tahap *pre-processing* data adalah proses *cleaning* yang mana pada proses ini data yang digunakan akan dibersihkan dari beberapa unsur

seperti angka dan juga simbol agar data yang digunakan untuk proses selanjutnya bisa lebih berkualitas. Contoh dari proses *cleaning* terdapat di Tabel 3.2.

Tabel 3. 2 Contoh dari proses *Cleaning*

Sebelum	Sesudah
Saya ingin lulus dengan IPK 4.0 di perkuliahan	Saya ingin lulus dengan IPK di perkuliahan

3.1.6 *Tokenizing*

Pada tahap *tokenizing*, data dipecah dari yang awalnya berbentuk kalimat atau paragraf menjadi kumpulan kata tunggal atau *term*. Contoh proses *tokenizing* terdapat di Tabel 3.3.

Tabel 3. 3 Contoh dari proses *Tokenizing*

Sebelum	Sesudah
Saya sedang belajar di sebuah Universitas di luar kota.	["Saya","sedang", "belajar", "di", "sebuah", "Universitas", "di", "luar", "kota"]

3.1.7 *Lemmatization*

Proses ini merupakan proses pengubahan kata menjadi ke bentuk dasarnya. Pada proses ini, setiap kata atau *term* yang ada akan dihilangkan imbuhan (*prefix/suffix*). Contoh dari proses *lemmatization* terdapat di Tabel 3.4.

Tabel 3. 4 Contoh dari proses *Lemmatization*

Sebelum	Sesudah
Saya sedang memperbaiki tugas yang telah diperiksa oleh dosen	Saya sedang perbaiki tugas yang telah periksa oleh dosen

3.1.8 *Penghapusan Stopwords*

Proses penghapusan *stopwords* ini merupakan proses pembersihan kumpulan kata yang ada di dalam daftar kata yang telah ditentukan atau *stoplist*. Proses ini dimaksudkan supaya data yang digunakan dalam penelitian tidak terdapat *common words* yang tidak bermakna seperti kata ganti subjek maupun kata hubung yang biasa digunakan dalam kalimat. Sehingga data yang digunakan sesuai dengan tujuan penelitian. Contoh proses *penghapusan stopwords* terdapat di Tabel 3.5.

Tabel 3. 5 Contoh dari proses Penghapusan *Stopwords*

Sebelum	Sesudah
Saya sedang belajar di sebuah Universitas di luar kota.	belajar Universitas luar kota

3.1.9 *Bag Of Words*

Selanjutnya, kerangka yang berisikan kumpulan kata dibentuk ke dalam sebuah model dengan menggunakan *Bag Of Words*. *Bag Of Words* berfungsi membentuk model dari kata di dalam dokumen dan menghitung jumlah kemunculan di tiap katanya. Nilai perhitungan frekuensi kemunculan tiap kata pada model *Bag Of Words* digunakan dalam proses Pemodelan Topik menggunakan LDA.

3.1.10 *Latent Dirichlet Allocation*

Setelah mendapatkan hasil matriks *term* dari proses *Bag Of Words*, dilakukan proses pemodelan topik LDA. Pemodelan topik LDA ini dapat memproses dokumen yang berukuran besar untuk mendapatkan informasi tersembunyi di dalamnya. Proses yang dilakukan di dalam pemodelan topik LDA adalah menentukan jumlah topik dan juga menentukan jumlah iterasi yang digunakan untuk mendapatkan hasil berupa topik-topik yang tersusun dari beberapa *term* yang saling berhubungan satu sama lainnya. Kemudian juga dapat dilakukan pemvisualisasian agar lebih mudah dalam menganalisis dan juga mengambil kesimpulan dari proses LDA yang telah dilakukan.

3.1.11 Pencarian Jumlah Topik

Sebelum melakukan analisis lanjutan (analisis tren topik), perlu dilakukan pencarian jumlah topik yang harus digunakan untuk model LDA. Metode *Topic coherence* merupakan salah satu dari banyak metode untuk membantu mencari jumlah topik untuk model akan digunakan pada *topic modeling*. Pada *topic coherence*, jumlah topik dalam suatu model bisa dicari dengan menghitung nilai *coherence* pada tiap topiknya. Semakin tinggi nilai *coherence* yang didapat maka semakin jelas model tersebut untuk diinterpretasikan oleh manusia dan pantas untuk digunakan dalam model LDA.

BAB IV

HASIL & PEMBAHASAN

4.1 Hasil Dari Penelitian

Berikut merupakan hasil proses-proses yang ada di dalam penelitian.

4.1.1 Hasil Pengumpulan Data

Data yang dikumpulkan menggunakan ekstensi webscrapper pada *web browser chrome*. Cara Kerja yang dilakukan pada ekstensi ini adalah dengan membuat selector graph sesuai dengan data yang dibutuhkan pada penelitian. Gambar 4.1 menunjukkan selector graph yang dibuat pada penelitian ini.



Gambar 4. 1 Selector Graph untuk proses *Scrapping* Abstrak

Pada proses pengumpulan data, selector yang diambil adalah abstrak artikel, bulan artikel, dan artikel list. Selector abstrak artikel berisi tentang abstrak-abstrak artikel yang dipilih, sedangkan selector bulan artikel berisi tentang tahun publish dari artikel tersebut. Selector artikel list berisi tentang judul dari artikel-artikel yang dikumpulkan.

Data yang diambil merupakan artikel-artikel yang terdapat pada jurnal-jurnal dengan *keyword* “Sistem Informasi/*Information System*” pada judul jurnal tersebut. Jurnal yang diambil sebagai data penelitian adalah jurnal-jurnal yang terindeks dengan nilai indeks Sinta 1 hingga Sinta 4. Total jumlah jurnal yang digunakan dalam penelitian ini yaitu 30 Jurnal. Tabel 4.1 menjabarkan jurnal-jurnal yang diambil sebagai data pada penelitian ini.

Tabel 4. 1 Indeks Jurnal dan Jurnal yang digunakan sebagai data penelitian

Indeks Sinta	Jurnal	Jumlah Data							
		2015	2016	2017	2018	2019	2020	2021	Sisa
Sinta 1	Register: Jurnal Ilmiah Teknologi Sistem Informasi	14	27	28	29	29	20	18	0

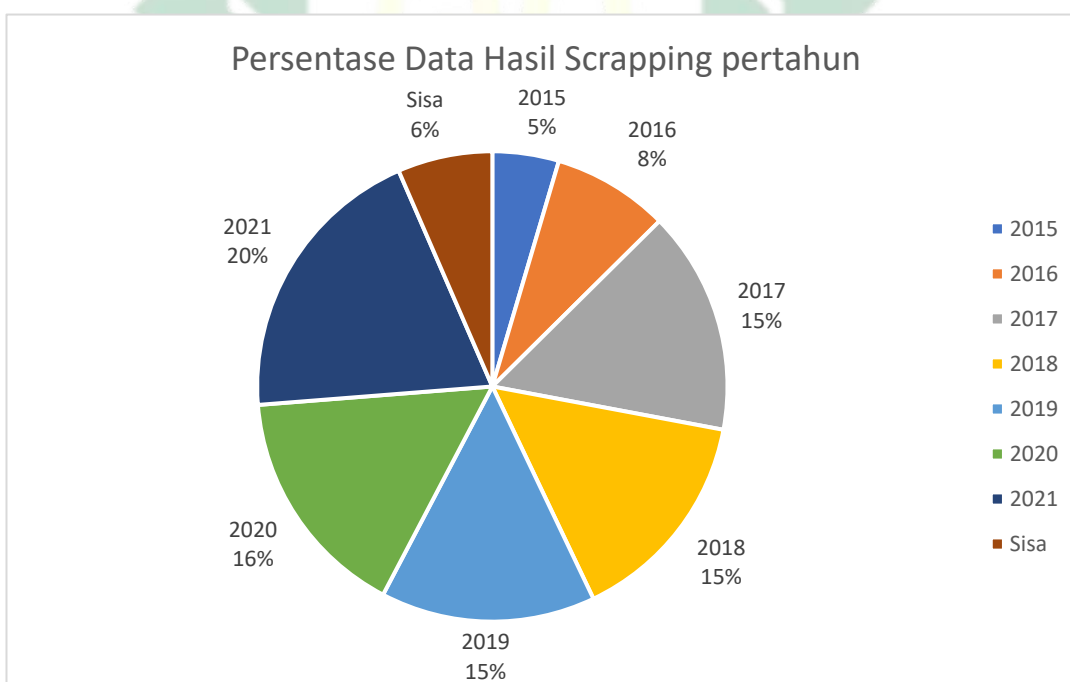
Sinta 2	INTENSIF: Jurnal Penelitian dan Penerapan Teknologi Sistem Informasi	0	0	24	16	16	20	20	30
	JSINBIS (Jurnal Sistem Informasi Bisnis)	20	19	25	28	31	28	14	22
	Jurnal Sistem Informasi (Journal of Information System)	24	13	13	11	0	0	0	0
	Journal of Information Systems Engineering and Business Intelligence	18	20	18	21	22	17	17	7
	Kinetik: Game Technology, Information System, Computer Network, Computing Electronics, and Control	0	36	91	99	42	44	39	9
Sinta 3	Jurnal Sisfokom (Sistem Informasi dan Komputer)	20	20	15	29	26	13	60	80
	Sistemasi: Jurnal Sistem Informasi	18	18	18	34	55	52	62	3
	JURTEKSI (Jurnal Teknologi dan Sistem Informasi)	0	0	15	14	50	41	45	5
	Jurnal Nasional Teknologi dan Sistem Informasi (TEKNOSI)	6	36	42	19	16	19	20	27

	Jurnal AKSI (Akuntansi dan Sistem Informasi)	0	0	18	12	16	21	103	6
	RESEARCH: Computer, Information System & Technology Management	0	0	0	12	12	12	18	6
	Pilar Nusa Mandiri: Journal of Computing and Information System	21	23	35	42	45	39	23	2
	Applied Information System and Management	0	0	0	20	0	10	38	2
Sinta 4	JSII (Jurnal Sistem Informasi)	0	0	72	32	40	19	23	4
	TEKNOLOGI: Jurnal Ilmiah Sistem Informasi	0	28	0	0	0	21	20	2
	Rabit: Jurnal Teknologi dan Sistem Informasi Univrab	0	12	14	14	12	14	18	6
	JuTISI (Jurnal Teknik Informatika dan Sistem Informasi)	38	41	48	43	26	69	108	7
	JUSIM (Jurnal Sistem Informasi Musirawas)	0	21	35	28	21	17	19	9
	JUSIFO: Jurnal Sistem Informasi	7	14	14	15	10	10	10	0

Jusikom: Jurnal Sistem Informasi Ilmu Komputer	0	0	6	20	22	16	11	9
Jurnal Teknologi Komputer dan Sistem Informasi	0	0	0	22	19	21	19	9
ZONAsi: Jurnal Sistem Informasi	0	0	0	0	15	23	13	9
PROCESSOR Jurnal Ilmiah Sistem Informasi, Teknologi Informasi dan Sistem Komputer	0	0	110	56	28	30	26	0
JOISIE (Journal Of Information Systems And Informatics Engineering)	0	0	0	6	41	20	17	6
JOURNAL OF INFORMATION SYSTEM RESEARCH (JOSH)	0	0	0	0	9	47	33	1
IJIS (Indonesian Journal on Information System)	0	9	11	10	9	18	18	5
Journal of Information System, Graphics, Hospitality and Technology	0	0	0	0	12	12	12	4
@is The Best: Accounting Information Systems and Information	0	0	6	12	6	18	12	6

Technology Business Enterprise									
SYSTEMIC: Information System and Informatics Journal	13	12	11	8	13	8	23	8	

Tabel 4.1 menampilkan hasil dari proses scrapping data di *website* *garuda.kemdikbud.go.id*. Dari proses tersebut, didapatkan data abstrak pada jurnal-jurnal seperti yang dipetakan didalam Tabel 4.1. Kolom Sisa merupakan jumlah data yang tidak termasuk ke dalam batas penelitian ini. Persentase data yang didapatkan dari hasil proses scrapping berjumlah 4354 baris dengan rincian pada Gambar 4.2 berikut.



Gambar 4. 2 Persentase Data Hasil Scrapping petahun

Pada setiap baris data terdapat informasi tentang judul dari artikel, tanggal dan tahun rilis artikel, dan abstrak dari artikel. Setelah mendapatkan data dari proses *scrapping*, maka selanjutnya akan dilakukan proses *pre-processing*.

4.1.2 Hasil Proses *Filtering* dan *Remove Duplicates*

Sebelum dilakukan proses *pre-processing*, data-data yang telah *discrapping* terlebih dahulu *filter* untuk mengumpulkan artikel-artikel yang terpublish di tahun yang sama agar menjadi sebuah corpus yang nantinya digunakan di proses-proses selanjutnya. Selain itu juga dilakukan proses *remove duplicates* untuk menghilangkan abstrak-abstrak yang muncul dua kali atau lebih sehingga hasil yang diperoleh di proses selanjutnya maksimal. Tabel 4.2 menjabarkan hasil yang didapat dari proses *filtering* dan *remove duplicates*.

Tabel 4. 2 Hasil dari proses *Filtering* dan *Remove Duplicates*

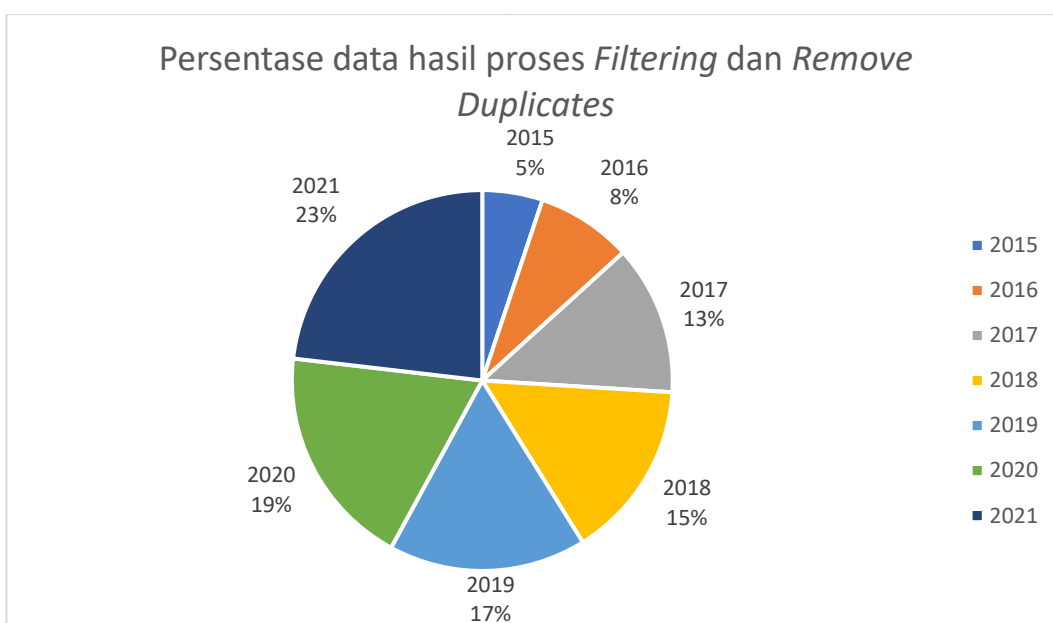
Indeks Sinta	Jurnal	Jumlah Data						
		2015	2016	2017	2018	2019	2020	2021
Sinta 1	Register: Jurnal Ilmiah	13	20	20	16	17	10	18
	Teknologi Sistem Informasi							
Sinta 2	INTENSIF: Jurnal Penelitian dan Penerapan Teknologi Sistem Informasi	0	0	17	16	16	20	20
	JSINBIS (Jurnal Sistem Informasi Bisnis)	19	19	22	28	31	27	14
	Jurnal Sistem Informasi (Journal of Information System)	22	6	12	10	0	0	0
	Journal of Information Systems Engineering and Business Intelligence	9	16	18	19	22	17	17
	Kinetik: Game Technology, Information System, Computer Network, Computing Electronics, and Control	0	17	43	56	36	43	38

Sinta 3	Jurnal Sisfokom (Sistem Informasi dan Komputer)	20	20	15	29	24	12	60
	Sistemasi: Jurnal Sistem Informasi	18	18	17	34	54	52	61
	JURTEKSI (Jurnal Teknologi dan Sistem Informasi)	0	0	15	14	50	41	45
	Jurnal Nasional Teknologi dan Sistem Informasi (TEKNOSI)	6	36	42	19	16	19	20
	Jurnal AKSI (Akuntansi dan Sistem Informasi)	0	0	0	7	16	21	73
	RESEARCH: Computer, Information System & Technology Management	0	0	0	12	12	12	18
	Pilar Nusa Mandiri: Journal of Computing and Information System	21	23	35	42	45	39	23
	Applied Information System and Management	0	0	0	20	0	10	38
Sinta 4	JSii (Jurnal Sistem Informasi)	0	0	34	18	26	19	23
	TEKNOLOGI: Jurnal Ilmiah Sistem Informasi	0	14	0	0	0	12	10
	Rabit: Jurnal Teknologi dan Sistem Informasi Univrab	0	12	14	14	12	13	18
	JuTISI (Jurnal Teknik Informatika dan Sistem Informasi)	34	41	48	43	26	69	108
	JUSIM (Jurnal Sistem Informasi Musirawas)	0	10	21	14	14	17	19

JUSIFO: Jurnal Sistem Informasi	7	14	14	15	10	10	10
Jusikom: Jurnal Sistem Informasi Ilmu Komputer	0	0	6	20	22	16	11
Jurnal Teknologi Komputer dan Sistem Informasi	0	0	0	22	19	21	19
ZONAsi: Jurnal Sistem Informasi	0	0	0	0	7	14	13
PROCESSOR Jurnal Ilmiah Sistem Informasi, Teknologi Informasi dan Sistem Komputer	0	0	28	32	28	30	26
JOISIE (Journal Of Information Systems And Informatics Engineering)	0	0	0	6	41	20	17
JOURNAL OF INFORMATION SYSTEM RESEARCH (JOSH)	0	0	0	0	9	47	33
IJIS (Indonesian Journal on Information System)	0	9	11	10	9	18	18
Journal of Information System, Graphics, Hospitality and Technology	0	0	0	0	12	12	12
@is The Best: Accounting Information Systems and Information Technology Business Enterprise	0	0	6	12	6	18	12

SYSTEMIC: Information System and Informatics Journal	12	12	11	7	13	8	23
--	----	----	----	---	----	---	----

Setelah melalui proses *filtering* dan *remove duplicates*, penelitian ini akan menggunakan data sejumlah 3529 abstrak artikel penelitian dengan rincian persentase pertahunnya ada pada Gambar 4.3 berikut.



Gambar 4. 3 Persentase data hasil proses *filtering* dan *remove duplicates*

4.1.3 Hasil *Pre-processing* Data

Proses yang dilakukan selanjutnya adalah proses *pre-processing* untuk membersihkan data yang telah *discrapping* dan juga diproses *filtering* serta *remove duplicates* dari *website* garuda.kemdikbud.go.id

4.1.3.1 Hasil Proses *Translating*

Pada tahap ini data-data artikel yang dikumpulkan masih dalam bentuk Bahasa yang campuran (Bahasa Indonesia dan Bahasa Inggris), oleh karena hal tersebut perlu dilakukan penyeragaman agar mudah saat melakukan proses selanjutnya. Data artikel diseragamkan menjadi Bahasa Inggris seperti yang tertera pada *pseudocode* berikut.

Translating
Deskripsi: <i>abstrak_artikel: data abstrak yang masih berbentuk raw (campuran bahasa indonesia dan inggris).</i> <i>translate: fungsi translate pada library translators</i>
<pre> START INPUT abtrak_artikel for i in abstrak_artikel do if (characters(i) != English) do translate.google(i); else do continue; END if END for OUTPUT abstrak_translation END </pre>

Pada proses *translating*, *library translators* digunakan untuk menerjemahkan data yang masih berbahasa Indonesia ke dalam Bahasa Inggris. Mesin penerjemah yang digunakan adalah *google* dikarenakan tidak semua mesin penerjemah menyediakan fitur terjemahan dari bahasa indonesia. Hasil proses *translating* terdapat di Tabel 4.3 berikut.

Tabel 4. 3 Hasil dari Proses *Translating*

<i>Sebelum Translating</i>	<i>Sesudah Translating</i>
<p><i>Abstrak- Penelitian ini bertujuan merancang dan membangun sistem pendukung keputusan untuk meramalkan jumlah kunjungan pasien RSUD Dr. Wahidin Sudiro Husodo Kota Mojokerto dengan menggunakan metode Extreme Learning Machine (ELM). Dengan adanya sistem pendukung keputusan ini direktur Rumah Sakit dapat meramalkan jumlah kunjungan pasien dan membantu dalam pembuatan kebijakan rumah sakit, mengatur sumber daya manusia dan keuangan, serta mendistribusikan sumber daya</i></p>	<p><i>ABSTRACT - This study aims to design and build a decision support system to predict the number of patients with Dr. Wahidin Sudiro Husodo Mojokerto City using the Extreme Learning Machine (ELM) method. With this decision support system the Director of Hospitals can predict the number of patient visits and assist in the making of hospital policies, regulates human and financial resources, and distribute material resources correctly, especially in dental poly. In the design of the decision support system</i></p>

<i>material dengan benar khususnya pada poli gigi. Dalam rancang bangun sistem pendukung keputusan ini dilakukan dalam beberapa tahap.</i>	<i>this decision is carried out in several stages.</i>
--	--

4.1.3.2 Hasil Proses Case Folding

Pada bagian ini, data hasil dari proses *cleaning*, dilakukan penyeragaman menjadi huruf kecil baik itu kata, angka dan tanda baca. *Pseudocode* dari proses *case folding* adalah sebagai berikut:

CASE FOLDING
Deskripsi: <i>Abstrak_translation: data abstrak hasil proses translasi</i> <i>Lower: fungsi lowering yang ada pada library string</i>
<i>START</i> <i>INPUT abtrak_translation</i> <i>for i in abstrak_translation do</i> <i>Lower(i)</i> <i>END for</i> <i>OUTPUT abstrak_lower</i> <i>END</i>

Pada proses *case folding*, *library str* dengan fungsinya yaitu fungsi *lower* digunakan untuk membuat semua huruf dalam data yang ada menjadi bentuk kecil semua. Hasil dari proses *case folding* ada di Tabel 4.4 berikut.

Tabel 4. 4 Hasil dari Proses Case Folding

<i>Sebelum Case folding</i>	<i>Sesudah Case folding</i>
<i>Employee recruitment is one of the activities that plays an important role in determining the quality of human resources in the company, so it is necessary to manage good employee recruitment data. The application of the database system is very helpful for the data collection process of employee recruitment, especially in terms of avoiding</i>	<i>employee recruitment is one of the activities that plays an important role in determining the quality of human resources in the company, so it is necessary to manage good employee recruitment data. the application of the database system is very helpful for the data collection process of employee recruitment, especially in terms of avoiding</i>

<i>human error, minimizing costs, and facilitating data management that can be continued into reports.</i>	<i>human error, minimizing costs, and facilitating data management that can be continued into reports.</i>
--	--

4.1.3.3 Hasil Proses *Cleaning*

Pada proses *Cleaning*, dilakukan pembersihan data dari unsur-unsur yang tidak memiliki arti maupun nilai di dalam data. Unsur seperti angka dan tanda baca dihilangkan agar data bersih dan bisa terfokus kepada kata-kata yang nantinya diproses. *Pseudocode* dari proses *cleaning* adalah sebagai berikut:

CLEANING
Deskripsi
<i>Abstrak_lower: data abstrak hasil dari proses case folding</i>
<i>START</i>
<i>INPUT abstrak_lower</i>
<i>for i in abstrak_lower do:</i>
<i>Remove number</i>
<i>Remove other symbol</i>
<i>END for</i>
<i>OUTPUT abstrak_clean</i>
<i>END</i>

Pada proses *cleaning*, *library* re dengan fungsi sub digunakan untuk membersihkan data dari objek-objek seperti angka, tanda baca dan *whitespace*. Hasil proses *cleaning* ada di Tabel 4.5 berikut

Tabel 4. 5 Hasil dari Proses *Cleaning*

<i>Sebelum Cleaning</i>	<i>Sesudah Cleaning</i>
<i>differences in service costs such as transfer fees and administrative costs adopted by several institutions raises several attitudes from customers, one of which is related to customer loyalty. loyalty is a behavior that is indicated by repeated use based on decision making. jenius mobile banking is a banking</i>	<i>differences in service costs such as transfer fees and administrative costs adopted by several institutions raises several attitudes from customers one of which is related to customer loyalty loyalty is behavior that is indicated by repeated use based on decision making jenius mobile banking is banking</i>

<p>service provided by the national pension savings bank (btpn). not yet known the factors that influence user loyalty, making genius mobile banking has not been maximized in product development, costs are also one of the factors that are not yet known so that the decision to change costs is unknown. the status of the loyalty of genius mobile banking users is also unknown, so the decision to make the loyalty program has not been carried out optimally.</p>	<p>service provided by the national pension savings bank btpn not yet known the factors that influence user loyalty making genius mobile banking has not been maximized in product development costs are also one of the factors that are not yet known so that the decision to change costs is unknown the status of the loyalty of genius mobile banking users is also unknown so the decision to make the loyalty program has not been carried out optimally</p>
---	---

4.1.3.4 Hasil Proses *Tokenizing*

Tokenizing dilakukan untuk memecah data yang awalnya berbentuk sebuah kalimat ke dalam bentuk kumpulan kata atau yang biasa disebut sebagai *token*. Dalam proses ini menggunakan *library python NLTK* yaitu fungsi *RegexTokenizer*. Berikut merupakan *pseudocode* dari proses *tokenizing*.

TOKENIZING
Deskripsi: <i>Abstrak_clean</i> : data abstrak hasil dari proses <i>cleaning</i> <i>RegexTokenizer</i> : fungsi <i>tokenizing</i> dalam <i>library nltk.tokenize</i>
<pre> START INPUT abstrak_clean for i in abstrak_clean do: RegexTokenizer(i) END for OUTPUT abstrak_token END </pre>

Pada proses *tokenizing*, *library nltk* dengan fungsi *RegexTokenizer* digunakan untuk mengubah data yang awalnya berbentuk kalimat/string menjadi ke bentuk token/kata. Hasil proses *tokenizing* ada di Tabel 4.6 berikut

Tabel 4. 6 Hasil dari proses *Tokenizing*

<i>Sebelum Tokenizing</i>	<i>Sesudah Tokenizing</i>
<p>health is an important factor in human life that has to be guarded both physical and mentally this study aimed to analyze the factors that affect health condition using medical check up data factors analyzed were consuming alcohol smoking exercise age and gender the method was the association rule using fpgrowth the result of this study was factors that affect the health condition is alcohol exercise and age this result evidenced by the rules which means that if person consumes more alcohol than days week with the amount of alcohol is less than ml day then health condition was poor with support and confidence which means that if one rarely exercise then health condition was poor with support and confidence which means that if person in middle age group then the condition of health was poor with support and confidence</p>	<p>[health, is, an, important, factor, in, human, life, that, has, to, be, guarded, both, physical, and, mentally, this, study, aimed, to, analyze, the, factors, that, affect, health, condition, using, medical, check, up, data, factors, analyzed, were, consuming, alcohol, smoking, exercise, age, and, gender, the, method, was, the, association, rule, using, fpgrowth, the, result, of, this, study, was, factors, that, affect, the, health, condition, is, alcohol, exercise, and, age, this, result, evidenced, by, the, rules, which, means, that, if, person, consumes, more, alcohol, than, days, week, with, the, amount, of, alcohol, is, less, than, ml, day, then, health, condition, was, poor, with, support, and, confidence, which, means, that, if, one, rarely, exercise, then, health, condition, was, poor, with, support, and, confidence, which, means, that, if, person, in, middle, age, group, then, the, condition, of, health, was, poor, with, support, and, confidence]</p>

4.1.3.5 Hasil Proses Lemmatization

Lemmatization yaitu proses mengembalikan bentuk suatu kata menjadi bentuk dasarnya. *Pseudocode* yang diterapkan pada proses *lemmatization* adalah sebagai berikut.

LEMMATIZATION
<p>Deskripsi <i>abstrak_token</i>: data abstrak hasil proses tokenizing <i>lemmatize</i>: fungsi lemmatize yang ada pada library WordNetLemmatizer</p>
<p>START</p> <pre> INPUT abstrak_token for i in abstrak_token do: if i != root word : Transform word to root word Else do: continue END for OUTPUT abstrak_lemma END </pre>

Pada proses *lemmatization*, data hasil dari proses *tokenizing* diubah menjadi bentuk dasarnya. Token-token yang berbentuk kata kerja, dan token-token yang menggunakan imbuhan (*prefiks*, *sufiks*, *infiks*, dan *confixes*) diubah menjadi bentuk kata dasarnya. Hasil proses *lemmatization* ada di Tabel 4.7 berikut.

Tabel 4.7 Hasil dari prsoes *Lemmatization*

Sebelum proses <i>Lemmatization</i>	Sesudah proses <i>Lemmatization</i>
<p>[in, current, business, practice, an, integration, between, business, and, it, is, important, enterprise, architecture, ea, is, one, of, the, studies, that, synergize, it, and, business, the, principle, will, be, use, as, the, main, base, to, design, appropriate, ea, before, describing, these, principles, it, is, necessary, to, understand, the, real, state, of, company, it, is, aims, to, make, the, situation, as, one,</p>	<p>[in, current, business, practice, an, integration, between, business, and, it, be, important, enterprise, architecture, ea, be, one, of, the, study, that, synergize, it, and, business, the, principle, will, be, use, as, the, main, base, to, design, appropriate, ea, before, describe, these, principles, it, be, necessary, to, understand, the, real, state, of, company, it, be, aim, to, make, the, situation, as, one,</p>

<p>as, of, benchmarks, to, describing, the, state, of, the, target, after, designing, state, of, the, target, it, will, make, processes, of, designing, ea, for, company, is, easier, this, paper, will, use, the, swot, analysis, method, to, recognize, their, core]</p>	<p>as, of, benchmarks, to, describe, the, state, of, the, target, after, design, state, of, the, target, it, will, make, process, of, design, ea, for, company, be, easier, this, paper, will, use, the, swot, analysis, method, to, recognize, their, core]</p>
--	--

4.1.3.6 Penghapusan *Stopwords*

Proses penghapusan *stopwords* merupakan proses penghapusan kata yang dianggap tidak memiliki makna dan dapat mempengaruhi hasil analisis. Pada proses ini digunakan *library python* NLTK untuk mendapatkan *stopwords* list yang sudah ada pada *library*. Terdapat juga beberapa *stopwords* tambahan yang ditambahkan karena tidak terdapat dalam list *stopwords* pada *library* NLTK seperti 'jsinbis', 'abstrak', 'abstrac', 'abstract', 'abstrack', 'background', 'intisari', 'abstraksi', 'ABSTRACTIVITIES', 'ABSTRUCERS', dan 'ea'. Penambahan kata tersebut dilakukan karena jumlah kemunculan kata-kata tersebut banyak di dalam data dan mempengaruhi hasil analisis. Berikut merupakan *pseudocode* yang digunakan dalam proses penghapusan *stopwords*:

PENGHAPUSAN STOPWORDS
<p>Deskripsi <i>abstrak_lemma</i>: Hasil dari proses Lemmatization <i>My_stopwords</i>: List Stopwords tambahan <i>Stopwords</i>: listss stopwords dari library NLTK ditambah dengan list <i>My_stopwords</i></p>
<pre> START INPUT abstrak_lemma for i in abstrak_lemma do: If i == word in stopwords: Remove i Else do: continue END for OUTPUT abstrak_poststopwords END </pre>

Pada proses penghapusan *stopwords*, *library* nltk digunakan untuk menghilangkan kata-kata yang tidak memiliki makna secara berarti. Hal tersebut dilakukan agar nantinya pada saat proses LDA dilakukan dapat menghasilkan topik yang berhubungan dengan penelitian. Di dalam proses ini *stopwords* yang digunakan adalah *stopwords* berbahasa Inggris dikarenakan data yang digunakan dalam Bahasa Inggris. Pada *library* nltk sudah terdapat kumpulan *stopwords* dalam bahasa Inggris. Selain hal tersebut, dapat juga dibuat *stopwords* tambahan dengan pertimbangan kata tersebut tidak berhubungan dengan topik yang sedang dibahas di dalam penelitian ini yaitu pembahasan tentang sistem informasi. Hasil dari proses penghapusan *stopwords* ada di Tabel 4.8 berikut.

Tabel 4. 8 Hasil dari proses penghapusan *Stopwords*

<i>Sebelum Proses Penghapusan Stopwords</i>	<i>Sesudah Proses Penghapusan Stopwords</i>
<p>[in, current, business, practice, an, integration, between, business, and, it, be, important, enterprise, architecture, ea, be, one, of, the, study, that, synergize, it, and, business, the, principle, will, be, use, as, the, main, base, to, design, appropriate, ea, before, describe, these, principles, it, be, necessary, to, understand, the, real, state, of, company, it, be, aim, to, make, the, situation, as, one, as, of, benchmarks, to, describe, the, state, of, the, target, after, design, state, of, the, target, it, will, make, process, of, design, ea, for, company, be, easier, this, paper, will, use, the, swot, analysis, method, to, recognize, their, core]</p>	<p>[current, business, practice, integration, business, enterprise, architecture, synergize, business, principle, main, principles, understand, real, state, company, aim, situation, benchmarks, state, target, state, target, company, easier, paper, swot, analysis, recognize, core]</p>

4.1.4 Hasil Proses Bag Of Words

Pada bagian *bag of words*, dilakukan penghitungan jumlah *term* atau kata yang terdapat di dalam data. Proses ini dilakukan untuk mendapatkan jumlah frekuensi kemunculan term/kata di dalam data tersebut. Hasil dari proses *bag of words* ini nantinya akan digunakan untuk proses LDA selanjutnya. Berikut merupakan *pseudocode* yang digunakan pada proses *Bag of Words*:

Bag of Words
Deskripsi
<i>abstrak_poststopwords</i> : Hasil dari proses penghapusan stopwords <i>dictionary</i> : fungsi <i>dictionary</i> yang ada pada library <i>gensim</i>
START
INPUT <i>abstrak_poststopwords</i> for <i>i</i> in <i>abstrak_poststopwords</i> do: count freq <i>dictionary</i> (<i>i</i>) END for OUTPUT <i>abstrak_bows</i>
END

Pada proses ini nantinya akan menghitung berapa jumlah kemunculan kata yang ada pada data di dalam dokumen. Berikut merupakan contoh hasil proses *Bag of Words* di Tabel 4.9.

Tabel 4. 9 Hasil dari proses *Bag of Words*

Hasil Proses Bag of Words
[('accord',1), ('aim',1), ('analysis',3), ('architecture',1), ('assessment',1), ('benchmarks',1), ('business',3), ('company',5), ('competencies',1), ('core',1), ('current',1), ('determine',2), ('direction',1), ('easier',1), ('enterprise',1), ('environment',1), ('foundations',1), ('integration',1), ('main',1), ('map',1), ('opportunities',1), ('order',1), ('organization',2), ('paper',1), ('position',4), ('practice',1), ('precise',1), ('principle',1), ('principles',3), ('quadrant',2), ('real',1), ('recognize',1), ('regions',1), ('resource',1), ('rsquo',1), ('situation',1), ('start',1), ('state',3), ('strategies',1), ('strategy',1), ('strengths',1), ('swot',3), ('synergize',1), ('target',3), ('threats',1), ('understand',1), ('weaknesses',1), ('wich',1)]

Berdasarkan Tabel 4.8 di atas, didapatkan bahwa pada dokumen yang ada di dalam data terdapat beberapa kata yang menyusunnya setelah melalui proses *pre-processing*. Contohnya kata ‘*accord*’ muncul sebanyak 1 kali dalam dokumen, kata ‘*analysis*’ muncul sebanyak 3 kali di dalam dokumen, dan kata ‘*position*’ muncul sebanyak 4 kali dalam dokumen.

4.1.5 Hasil Proses LDA

Pada proses LDA, dimulai dengan melakukan pembentukan model dari LDA yang nantinya akan digunakan pada penelitian. Setelah itu dilakukan evaluasi untuk menentukan jumlah topik yang pas untuk dilakukan proses LDA pada data yang ada di penelitian. Berikut merupakan *pseudocode* pada proses LDA.

Latent Dirichlet Allocation
<p>Deskripsi <i>Abstrak_bows</i>: Hasil proses Bag of Words <i>Topic</i>: jumlah topik yang digunakan dalam model <i>LdaModel</i>: fungsi untuk membentuk model LDA pada <i>gensim</i></p>
<pre> START INPUT abstrak_bows, topic, passes for j in topic_list do: LdaModel(i, abstrak_bows) return topic_lda END for OUTPUT topic_lda END </pre>

Pada proses LDA, fungsi *LdaModel* pada *library gensim* digunakan untuk mendapatkan topik-topik yang ada pada data penelitian. Setelah melakukan proses LDA, hasil belum bisa dianalisis dan disimpulkan karena masih perlu dilakukan evaluasi untuk menentukan jumlah topik yang pas untuk digunakan ke dalam model LDA yang telah dibuat sebelumnya. Metode *topic coherence* digunakan dalam penelitian ini untuk mendapatkan jumlah topik yang akan digunakan di dalam model LDA. Berikut merupakan *pseudocode* dari proses *Topic Coherence*:

Topic Coherence
<p>Deskripsi <i>Topic_list</i>: jumlah topik yang ingin diuji <i>ldamodel</i>: model lda yang digunakan sebagai objek pengujian <i>LdaModel</i>: fungsi untuk membentuk model LDA pada <i>gensim</i> <i>CoherenceModel</i>: fungsi untuk menguji model LDA\ <i>coherence</i>: jenis coherence test yang digunakan</p>
<pre> START </pre>


```

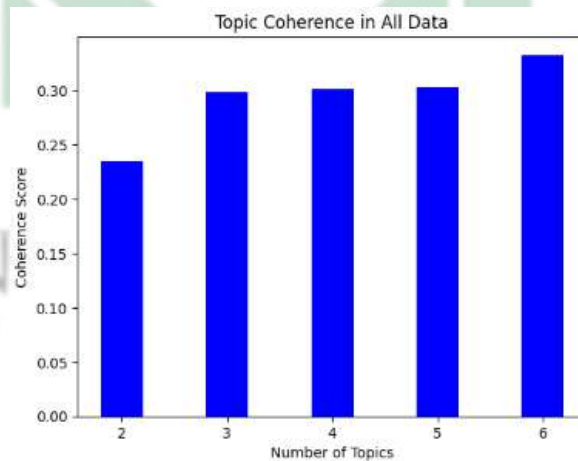
INPUT topic_list, passes, coherence
for i in LdaModel do:
    for j in topic_list do:
        LdaModel(i, ldamodel)
        return ldamodel
    END for
CoherenceModel(ldamodel, coherence)
return coherence_scores
END for

OUTPUT coherence_scores

END

```

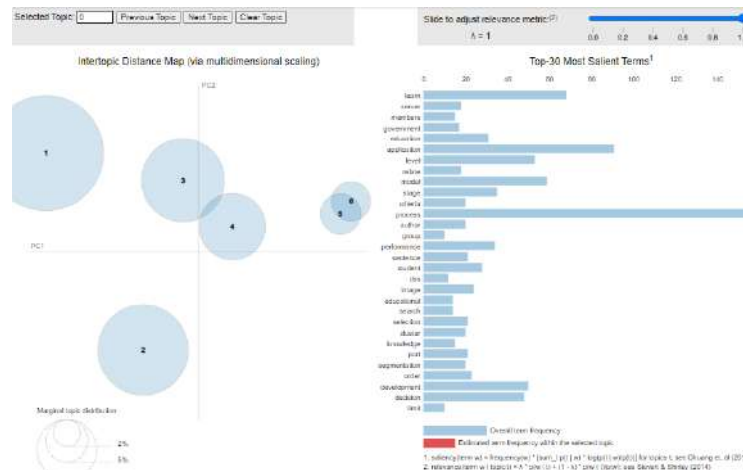
Pada proses *topic coherence*, fungsi *CoherenceModel* digunakan untuk mendapatkan *coherence score* yang nantinya digunakan untuk menentukan berapa jumlah topik yang pas digunakan ke dalam model LDA yang telah dibuat sebelumnya. Nilai *coherence* yang paling besar mengindikasikan bahwa jumlah topik tersebut pas digunakan untuk model LDA tersebut. Pada penelitian ini jumlah topik dicari dengan menggunakan data gabungan dari seluruh *corpus*. Jumlah topik yang akan dicari meliputi jumlah topik 2 hingga 10. Hasil dari metode *topic coherence* didapatkan bahwa jumlah topik 6 menjadi jumlah topik dengan nilai *coherence* terbesar.



Gambar 4. 4 Hasil *Topic Coherence*

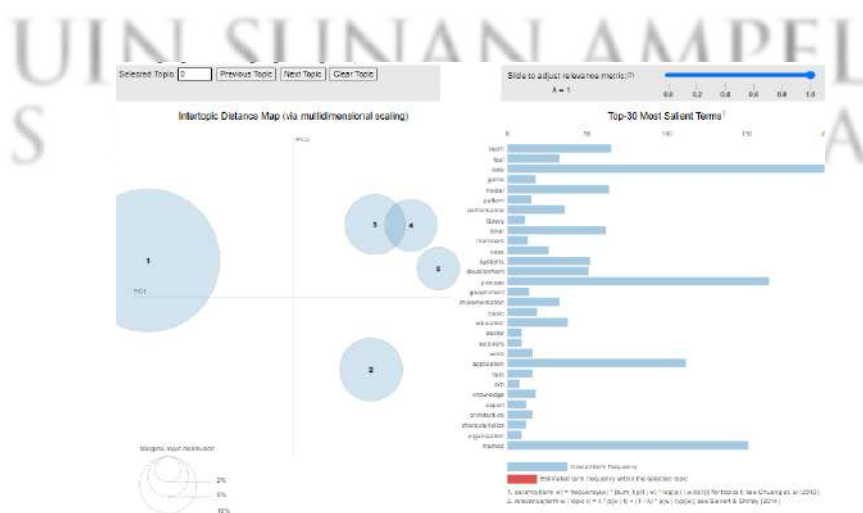
Namun, setelah hasil jumlah topik dari metode *topic coherence* tersebut diterapkan ke dalam model LDA yang sudah dibuat sebelumnya, topik-topik yang dihasilkan belum terlihat optimal. Hal tersebut dikarenakan masih banyaknya topik-topik yang beririsan satu sama lain dan juga berdekatan satu sama lain dalam artian kata-kata yang membentuk beberapa topik tersebut masih dalam lingkup yang sama

sehingga masih dapat dikatakan satu kesatuan topik. Hal tersebut nantinya akan mempengaruhi penginterpretasian tiap-tiap topik pada *corpus*. Gambar 4. 2 merupakan hasil pemetaan topik-topik LDA dengan jumlah topik 8.



Gambar 4. 5 Pemetaan topik LDA dengan jumlah topik 6

Pada dasarnya, semakin berjauhan topik saat divisualisasikan, maka semakin optimal model topik tersebut. Oleh karena hal tersebut, pada proses pencarian jumlah topik untuk model LDA, dilakukan pengurangan hingga jumlah topik menjadi optimal. Dengan mengacu pada hasil metode *topic coherence* yaitu jumlah topik 6, maka dilakukan pengurangan jumlah topik yang digunakan pada model kemudian divisualisasikan untuk membantu menentukan jumlah topik yang optimal. Gambar 4.6 merupakan gambar pemetaan topik LDA dengan jumlah topik 5.



Gambar 4. 6 Pemetaan topik LDA dengan jumlah topik 5

Pada Gambar 4.8 menampilkan bahwa jumlah topik 3 untuk digunakan pada model LDA di penelitian ini merupakan jumlah topik yang optimal. Hal tersebut dibuktikan dengan pemetaan pada Gambar 4.8 yang mana dari ketiga topik yang dihasilkan, ketiga topik tersebut tidak saling beririsan dan juga berdekatan yang berarti kata-kata yang merepresentasikan ketiga topik tersebut tidak saling berhubungan. Hasil dari proses pencarian jumlah topik menghasilkan jumlah topik 3 untuk digunakan dalam model LDA di penelitian ini.

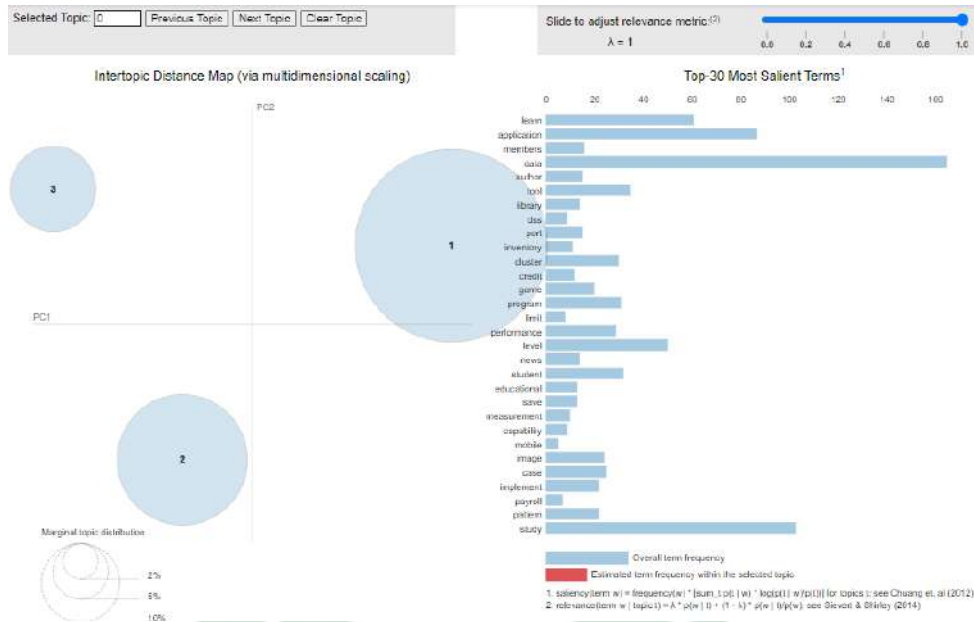
4.1.5.1 Hasil Proses LDA Data tahun 2015

Dengan menggunakan data abstrak artikel yang dipublikasikan pada tahun 2015, dengan menggunakan jumlah topik 3 maka topik-topik yang dihasilkan oleh proses LDA serta kata-kata penyusunnya ada pada Tabel 4.10 berikut

Tabel 4. 10 Hasil Proses LDA Data tahun 2015

Topic ID	Kata Penyusun
0	0.010*"learn" + 0.007*"process" + 0.006*"study" + 0.005*"tool" + 0.004*"level" + 0.004*"method" + 0.004*"design" + 0.004*"members" + 0.004*"performance" + 0.003*"test"
1	0.006*"application" + 0.005*"data" + 0.004*"process" + 0.004*"author" + 0.004*"study" + 0.003*"dss" + 0.003*"program" + 0.002*"inventory" + 0.002*"method" + 0.002*"level"
2	0.015*"data" + 0.012*"process" + 0.011*"method" + 0.007*"application" + 0.007*"study" + 0.006*"model" + 0.006*"design" + 0.005*"technology" + 0.005*"test" + 0.005*"support"

Sesuai dengan yang tertera pada Tabel 4.10, maka visualisasi yang dihasilkan dari proses LDA dengan *corpus* 2015 ada di Gambar 4.9 berikut.



Gambar 4. 9 Hasil visualisasi proses LDA corpus 2015

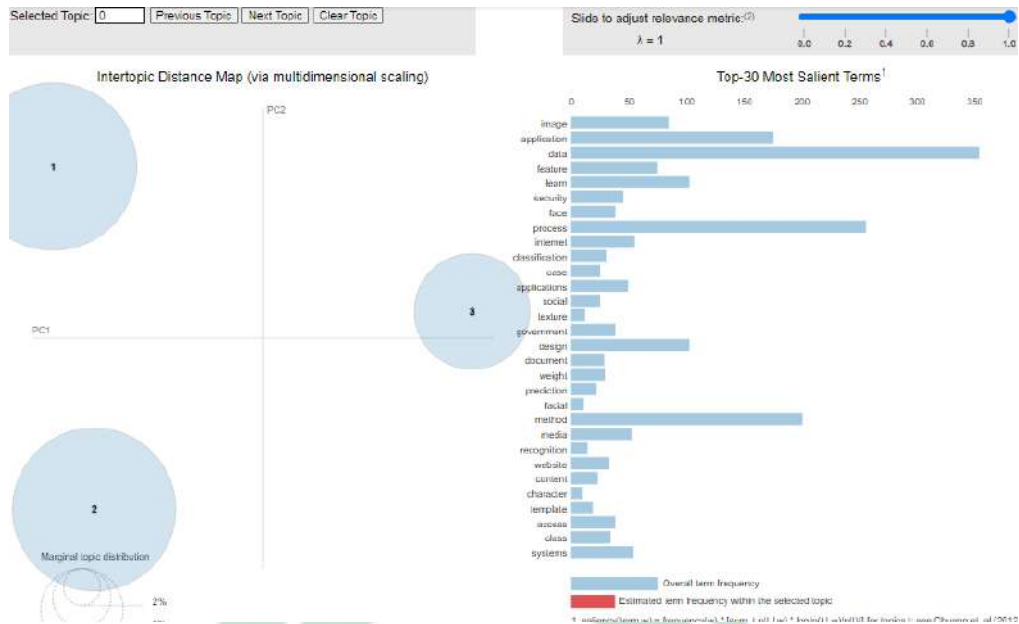
4.1.5.2 Hasil Proses LDA Data tahun 2016

Dengan menggunakan data abstrak artikel yang dipublikasikan pada tahun 2016, dengan menggunakan jumlah topik 3, maka topik-topik yang dihasilkan oleh proses LDA serta kata-kata penyusunnya ada pada Tabel 4.11 berikut

Tabel 4. 11 Hasil Proses LDA Data tahun 2016

Topic ID	Kata Penyusun
0	$0.012 * "image" + 0.012 * "process" + 0.009 * "method" +$ $0.007 * "data" + 0.006 * "feature" + 0.005 * "study" +$ $0.004 * "security" + 0.004 * "face" +$ $0.003 * "classification" + 0.003 * "decision"$
1	$0.014 * "application" + 0.008 * "process" +$ $0.008 * "method" + 0.007 * "learn" + 0.007 * "data" +$ $0.006 * "study" + 0.005 * "service" + 0.005 * "internet" +$ $0.005 * "time" + 0.004 * "performance"$
2	$0.024 * "data" + 0.011 * "process" + 0.008 * "test" +$ $0.007 * "method" + 0.006 * "service" + 0.006 * "design" +$ $0.006 * "development" + 0.006 * "study" +$ $0.005 * "technology" + 0.004 * "support"$

Sesuai dengan yang tertera pada Tabel 4.11, maka visualisasi yang dihasilkan dari proses LDA dengan corpus 2016 ada di Gambar 4.10 berikut.



Gambar 4. 10 Hasil visualisasi proses LDA *corpus* 2016

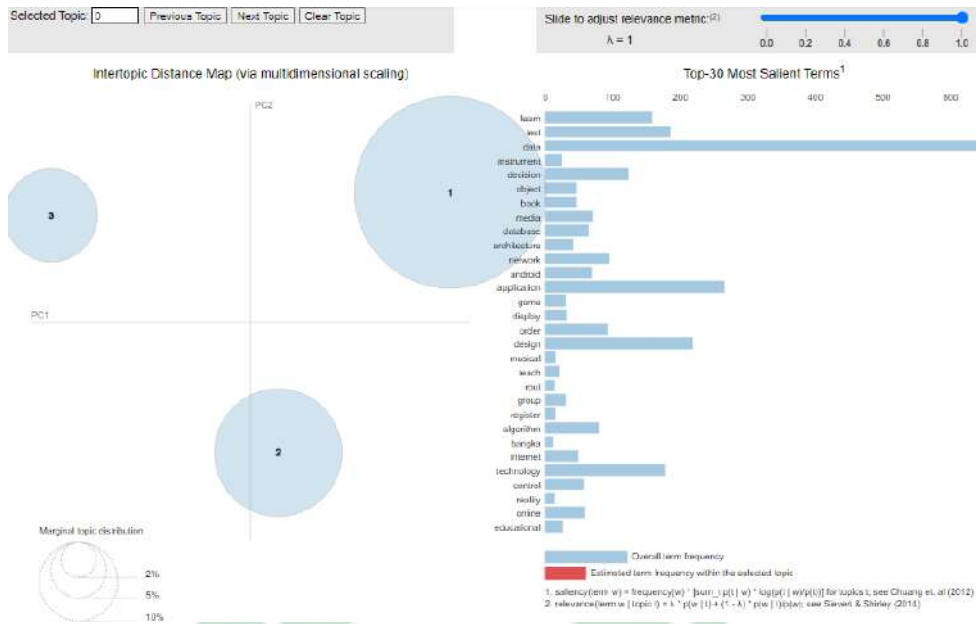
4.1.5.3 Hasil Proses LDA Data tahun 2017

Dengan menggunakan data abstrak artikel yang dipublikasikan pada tahun 2017, dengan menggunakan jumlah topik 3, maka topik-topik yang dihasilkan oleh proses LDA serta kata-kata penyusunnya ada di Tabel 4.12 berikut

Tabel 4. 12 Hasil Proses LDA Data tahun 2017

Topic ID	Kata Penyusun
0	0.013*"learn" + 0.008*"test" + 0.005*"application" + 0.004*"technology" + 0.004*"network" + 0.004*"object" + 0.004*"method" + 0.004*"instrument" + 0.004*"design" + 0.003*"game"
1	0.020*"data" + 0.015*"process" + 0.014*"method" + 0.009*"service" + 0.007*"study" + 0.007*"application" + 0.006*"support" + 0.006*"model" + 0.006*"technology" + 0.005*"design"
2	0.018*"data" + 0.009*"application" + 0.007*"design" + 0.007*"method" + 0.007*"process" + 0.006*"study" + 0.004*"android" + 0.004*"database" + 0.004*"book" + 0.004*"order"

Sesuai dengan yang tertera pada Tabel 4.12, maka visualisasi yang dihasilkan dari proses LDA dengan *corpus* 2017 ada di Gambar 4.11 berikut.



Gambar 4. 11 Hasil visualisasi proses LDA corpus 2017

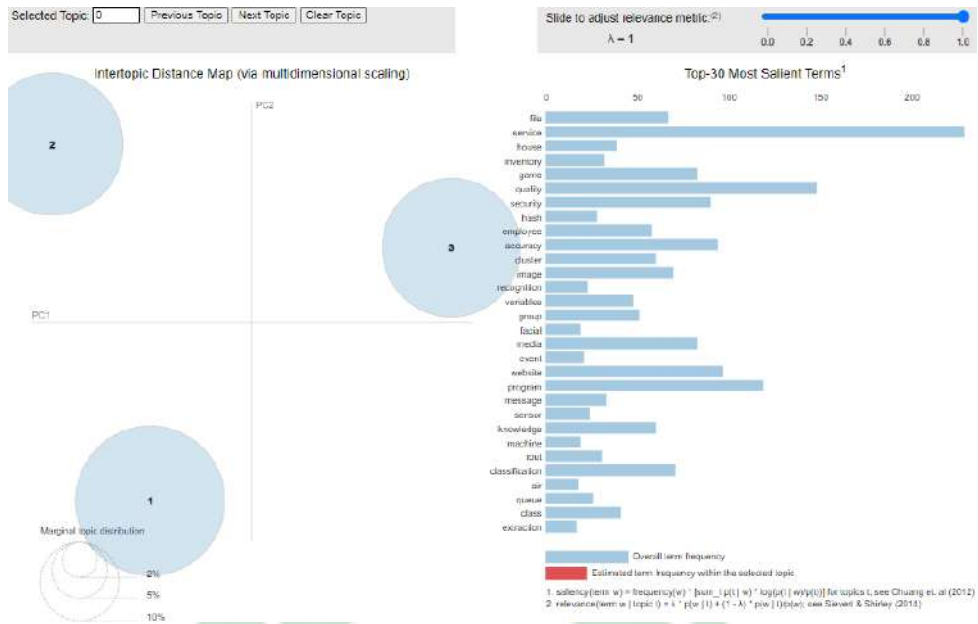
4.1.5.4 Hasil Proses LDA Data tahun 2018

Dengan menggunakan data abstrak artikel yang dipublikasikan pada tahun 2018, dengan menggunakan jumlah topik 3, maka topik-topik yang dihasilkan oleh proses LDA serta kata-kata penyusunnya ada pada Tabel 4.13 berikut.

Tabel 4. 13 Hasil Proses LDA Data tahun 2018

Topic ID	Kata Penyusun
0	0.017*"data" + 0.012*"method" + 0.010*"process" + 0.009*"application" + 0.008*"study" + 0.006*"technology" + 0.006*"design" + 0.006*"development" + 0.006*"service" + 0.005*"program"
1	0.011*"data" + 0.010*"process" + 0.009*"service" + 0.009*"method" + 0.008*"application" + 0.007*"study" + 0.006*"quality" + 0.006*"development" + 0.005*"technology" + 0.005*"test"
2	0.015*"data" + 0.010*"process" + 0.009*"method" + 0.008*"study" + 0.007*"model" + 0.006*"test" + 0.005*"design" + 0.005*"application" + 0.004*"time" + 0.004*"algorithm"

Sesuai dengan yang tertera pada Tabel 4.13, maka visualisasi yang dihasilkan dari proses LDA dengan corpus 2018 ada di Gambar 4.12 berikut.



Gambar 4. 12 Hasil visualisasi proses LDA corpus 2018

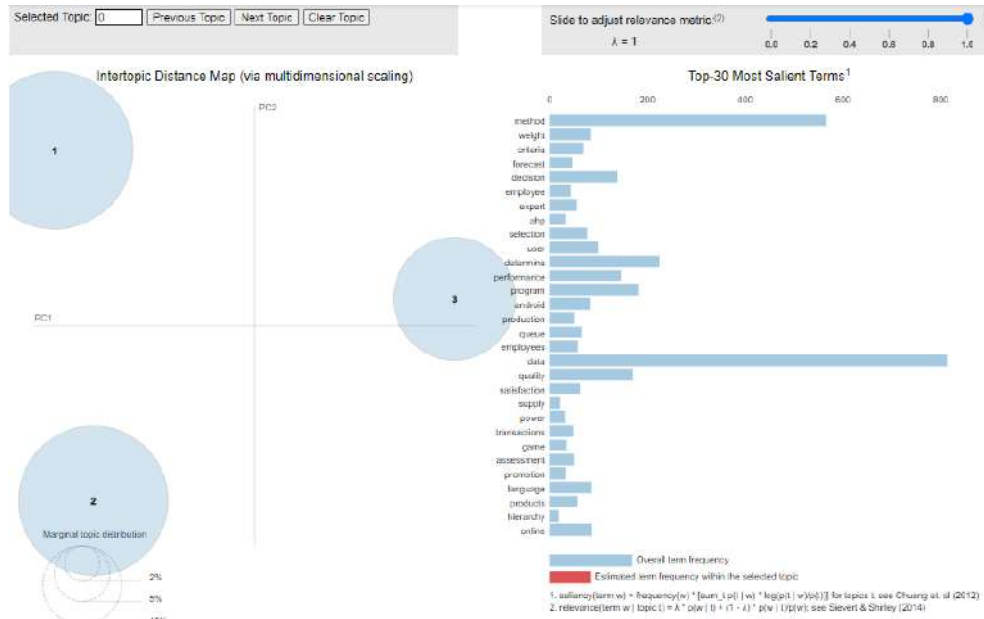
4.1.5.5 Hasil Proses LDA Data tahun 2019

Dengan menggunakan data abstrak artikel yang dipublikasikan pada tahun 2019, dengan menggunakan jumlah topik 3, maka topik-topik yang dihasilkan oleh proses LDA serta kata-kata penyusunnya ada di Tabel 4.14 berikut.

Tabel 4. 14 Hasil Proses LDA Data tahun 2019

Topic ID	Kata Penyusun
0	0.015*"data" + 0.012*"method" + 0.010*"process" + 0.010*"application" + 0.008*"design" + 0.008*"study" + 0.007*"service" + 0.007*"program" + 0.007*"development" + 0.007*"test"
1	0.022*"data" + 0.013*"process" + 0.011*"study" + 0.009*"service" + 0.008*"application" + 0.006*"method" + 0.006*"technology" + 0.005*"model" + 0.005*"report" + 0.005*"design"
2	0.018*"method" + 0.010*"process" + 0.008*"data" + 0.008*"determine" + 0.007*"study" + 0.006*"weight" + 0.006*"performance" + 0.005*"decision" + 0.005*"quality" + 0.005*"criteria"

Sesuai dengan yang tertera pada Tabel 4.14, maka visualisasi yang dihasilkan dari proses LDA dengan corpus 2019 ada di Gambar 4.13 berikut.



Gambar 4. 13 Hasil visualisasi proses LDA corpus 2019

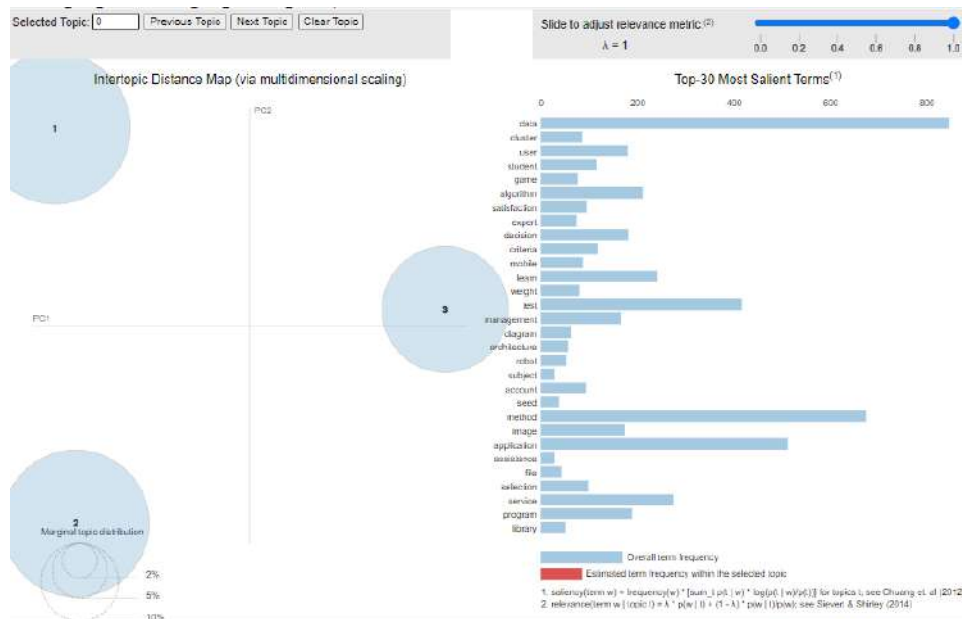
4.1.5.6 Hasil Proses LDA Data tahun 2020

Dengan menggunakan data abstrak artikel yang dipublikasikan pada tahun 2020, dengan menggunakan jumlah topik 3, maka topik-topik yang dihasilkan oleh proses LDA serta kata-kata penyusunnya ada di Tabel 4.15 berikut

Tabel 4. 15 Hasil Proses LDA Data tahun 2020

Topic ID	Kata Penyusun
0	0.013*"data" + 0.013*"method" + 0.009*"process" + 0.009*"study" + 0.007*"test" + 0.007*"algorithm" + 0.007*"learn" + 0.006*"application" + 0.006*"determine" + 0.005*"decision"
1	0.013*"application" + 0.013*"method" + 0.012*"study" + 0.009*"process" + 0.007*"service" + 0.007*"user" + 0.006*"design" + 0.006*"data" + 0.005*"analysis" + 0.005*"technology"
2	0.022*"data" + 0.012*"process" + 0.011*"study" + 0.010*"test" + 0.008*"method" + 0.007*"model" + 0.007*"analysis" + 0.007*"technology" + 0.006*"development" + 0.006*"design"

Sesuai dengan yang tertera pada Tabel 4.15, maka visualisasi yang dihasilkan dari proses LDA dengan corpus 2020 ada di Gambar 4.14 berikut.



Gambar 4. 14 Hasil visualisasi proses LDA *corpus* 2020

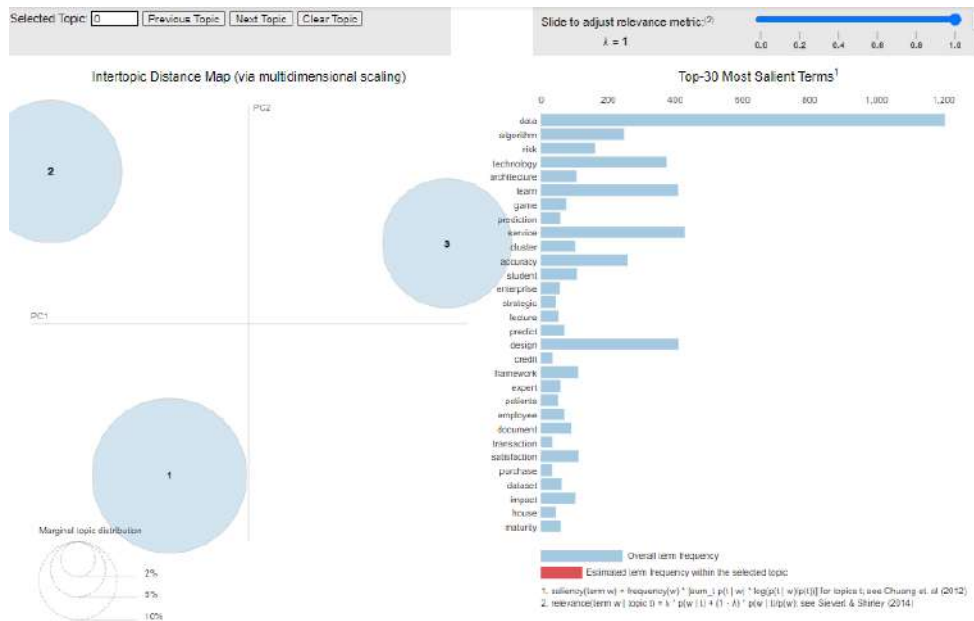
4.1.5.7 Hasil Proses LDA Data tahun 2021

Dengan menggunakan data abstrak artikel yang dipublikasikan pada tahun 2021, dengan menggunakan jumlah topik 3, maka topik-topik yang dihasilkan oleh proses LDA serta kata-kata penyusunnya ada pada Tabel 4.16 berikut

Tabel 4. 16 Hasil Proses LDA Data tahun 2021

Topic ID	Kata Penyusun
0	0.016*"study" + 0.013*"method" + 0.010*"application" + 0.010*"learn" + 0.010*"process" + 0.010*"data" + 0.008*"test" + 0.006*"management" + 0.005*"model" + 0.005*"risk"
1	0.035*"data" + 0.010*"study" + 0.010*"method" + 0.010*"process" + 0.008*"algorithm" + 0.007*"test" + 0.006*"accuracy" + 0.006*"model" + 0.005*"time" + 0.005*"application"
2	0.010*"process" + 0.010*"service" + 0.009*"technology" + 0.009*"method" + 0.009*"design" + 0.009*"application" + 0.008*"study" + 0.008*"development" + 0.007*"data" + 0.006*"model"

Sesuai dengan yang tertera pada Tabel 4.16, maka visualisasi yang dihasilkan dari proses LDA dengan *corpus* 2021 ada di Gambar 4.15 berikut.



Gambar 4. 15 Hasil visualisasi proses LDA corpus 2021

4.2 Pembahasan Analisis

Untuk dapat melakukan proses analisis tren topik, maka perlu dilakukan interpretasi topik-topik yang telah dihasilkan oleh proses LDA sebelumnya. Dengan berkonsultasi pada para ahli/dosen yang ada di bidang Sistem Informasi, maka ditentukan pada penelitian ini untuk merepresentasikan topik dengan keilmuan Teknologi (meliputi *Software Development*, dsb.), Sistem Cerdas (meliputi *Data Analyst*, dsb.), dan Manajemen (meliputi *IT Governance*, dsb.) sebagai representasi ketiga topik yang ada di tiap corpusnya. Untuk mendapatkan topik yang merepresentasikan tiap dokumen, penelitian ini menggunakan fungsi `get_document_topics()` pada `library gensim` untuk mendapatkan nilai probabilitas topik terhadap tiap dokumen. Tiap dokumen memungkinkan untuk mempunyai lebih dari 1 representasi topik, oleh karena itu pada penelitian ini mengambil topik dengan nilai probabilitas terbesar di tiap dokumennya sebagai representasi topik terhadap dokumen tersebut.

4.2.1 Analisis Topik pada tahun 2015

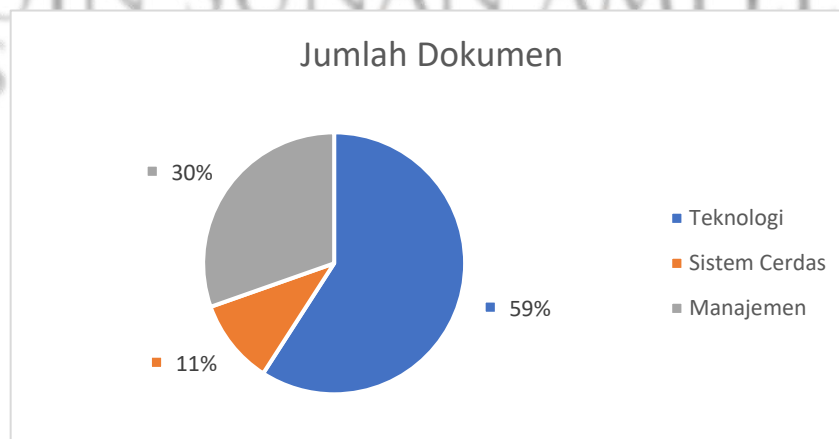
Dengan mengacu pada Tabel Hasil Proses LDA tahun 2015, topik-topik yang dihasilkan pada tahun 2015 masih berbentuk topik id. Oleh karena hal tersebut, topik-topik tersebut perlu diinterpretasikan untuk memudahkan dalam proses untuk mendapatkan tren pada jurnal lingkup Sistem Informasi. Cara penginterpretasian

yang dilakukan adalah dengan cara manual dengan konsultasi kepada ahli-ahli di bidang Sistem Informasi. Tabel 4.17 merupakan hasil interpretasi serta jumlah dokumen dari topik-topik yang ada di tahun 2015.

Tabel 4. 17 Hasil interpretasi topik tahun 2015

<i>Topic id</i>	<i>Kata Penyusun (top 10 kata)</i>	<i>Hasil Interpretasi</i>	<i>Jumlah dokumen</i>
0	"learn", "process", "study", "tool", "level", "method", "design", "members", "performance", "test"	Manajemen	55
1	"application", "data", "process", "author", "study", "dss", "program", "inventory", "method", "level"	Sistem Cerdas	19
2	"data", "process", "method", "application", "study", "model", "design", "technology", "test", "support"	Teknologi	107

Dengan mengacu pada Tabel 4.17 yang menampilkan jumlah dokumen terkait ke tiap topik, maka persentase jumlah dokumen terhadap topik di 2015 ada pada Gambar 4.16 berikut.



Gambar 4. 16 Persentase jumlah dokumen terhadap topik di 2015

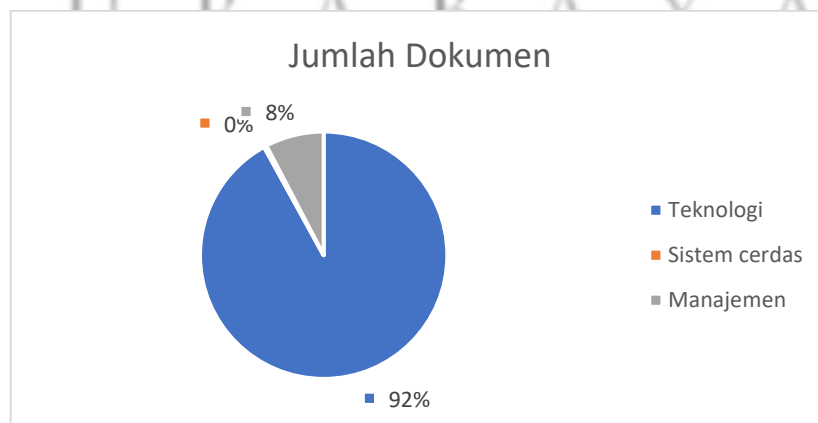
4.2.2 Analisis Topik pada tahun 2016

Pada hasil proses LDA tahun 2016 juga dilakukan interpretasi topik secara manual dengan mengacu pada tabel hasil proses LDA tahun 2016. Hal ini dilakukan untuk membantu proses analisis tren jurnal sistem informasi di tahun 2016. Tabel 4.18 merupakan tabel hasil dari interpretasi topik-topik yang ada di tahun 2016.

Tabel 4. 18 Hasil interpretasi topik tahun 2016

Topic id	Kata Penyusun (Top 10 kata)	Hasil Interpretasi	Jumlah Dokumen
0	"image", "process", "method", "data", "feature", "study", "security", "face", "classification", "decision"	Manajemen	22
1	"application", "process", "method", "learn", "data", "study", "service", "internet", "time", "performance"	Sistem Cerdas	1
2	"data", "process", "test", "method", "service", "design", "development", "study", "technology", "support"	Teknologi	264

Dengan mengacu pada Tabel 4.18 yang menampilkan jumlah dokumen terkait ke tiap topik, maka persentase jumlah dokumen terhadap topik di 2016 ada pada Gambar 4.17 berikut.



Gambar 4. 17 Persentase jumlah dokumen terhadap topik di 2016

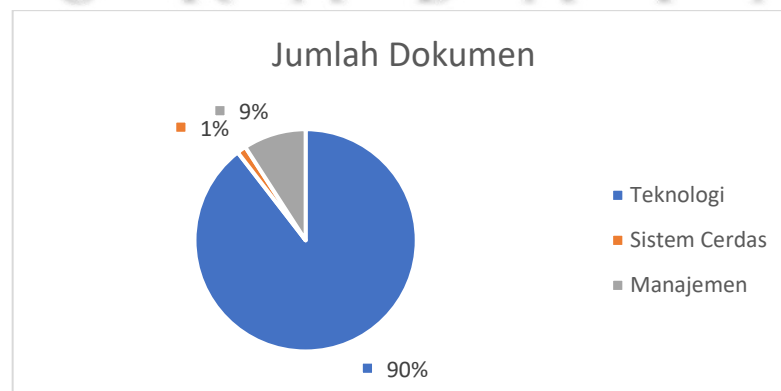
4.2.3 Analisis Topik pada tahun 2017

Pada hasil proses LDA tahun 2017 juga dilakukan penginterpretasian topik seperti tahun-tahun sebelumnya. Tabel 4.19 adalah tabel hasil dari interpretasi topik-topik pada tahun 2017.

Tabel 4. 19 Hasil interpretasi topik tahun 2017

Topic id	Kata Penyusun (Top 10 kata)	Hasil Interpretasi	Jumlah Dokumen
0	"learn", "test", "application", "technology", "network", "object", "method", "instrument", "design", "game"	Manajemen	41
1	"data", "process", "method", "service", "study", "application", "support", "model", "technology", "design"	Sistem Cerdas	6
2	"data", "application", "design", "method", "process", "study", "android", "database", "book", "order"	Teknologi	402

Dengan mengacu pada Tabel 4.19 yang menampilkan jumlah dokumen terkait ke tiap topik, maka persentase jumlah dokumen terhadap topik di 2017 ada pada Gambar 4.18 berikut



Gambar 4. 18 Persentase jumlah dokumen terhadap topik di 2017

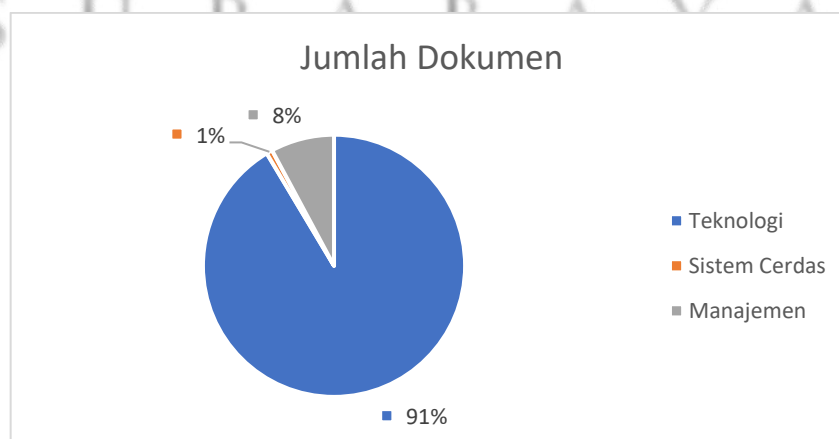
4.2.4 Analisis Topik pada tahun 2018

Seperti tahun-tahun sebelumnya, hasil proses LDA tahun 2018 juga dilakukan interpretasi topik-topiknya. Tabel 4.20 adalah tabel hasil dari interpretasi topik-topik pada tahun 2018.

Tabel 4. 20 Hasil interpretasi topik tahun 2018

<i>Topic id</i>	<i>Kata Penyusun (Top 10 kata)</i>	<i>Hasil Interpretasi</i>	<i>Jumlah Dokumen</i>
0	"data", "method", "process", "application", "study", "technology", "design", "development", "service", "program"	Manajemen	42
1	"data", "process", "service", *"method", "application", "study", "quality", "development", "technology", "test"	Sistem Cerdas	4
2	*"data", "process", "method", "study", "model", "test", "design", "application", "time", "algorithm"	Teknologi	490

Dengan mengacu pada Tabel 4.20 yang menampilkan jumlah dokumen terkait ke tiap topik, maka persentase jumlah dokumen terhadap topik di 2018 ada pada Gambar 4.19 berikut.



Gambar 4. 19 Persentase jumlah dokumen terhadap topik di 2018

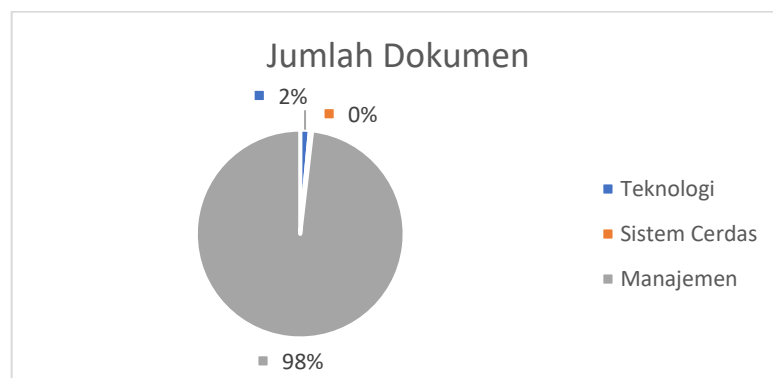
4.2.5 Analisis Topik pada tahun 2019

Pada hasil proses LDA tahun 2019 juga dilakukan interpretasi topik yang mengacu pada tabel hasil proses LDA tahun 2019. Hal ini dilakukan untuk membantu proses analisis tren jurnal sistem informasi di tahun 2019. Tabel 4.21 merupakan tabel hasil dari interpretasi topik-topik yang ada di tahun 2019.

Tabel 4. 21 Hasil interpretasi topik tahun 2019

<i>Topic id</i>	<i>Kata Penyusun (Top 10 kata)</i>	<i>Hasil Interpretasi</i>	<i>Jumlah Dokumen</i>
0	"data", "method", "process", "application", "design", "study", "service", "program", "development", "test"	Teknologi	46
1	"data", "process", "study", "service", "application", "method", "technology", "model", "report", "design"	Sistem Cerdas	2
2	"method", "process", "data", "determine", "study", "weight", "performance", "decision", "quality", "criteria"	Manajemen	545

Dengan mengacu pada Tabel 4.21 yang menampilkan jumlah dokumen terkait ke tiap topik, maka persentase jumlah dokumen terhadap topik di 2019 ada pada Gambar 4.20 berikut.



Gambar 4. 20 Persentase jumlah dokumen terhadap topik di 2019

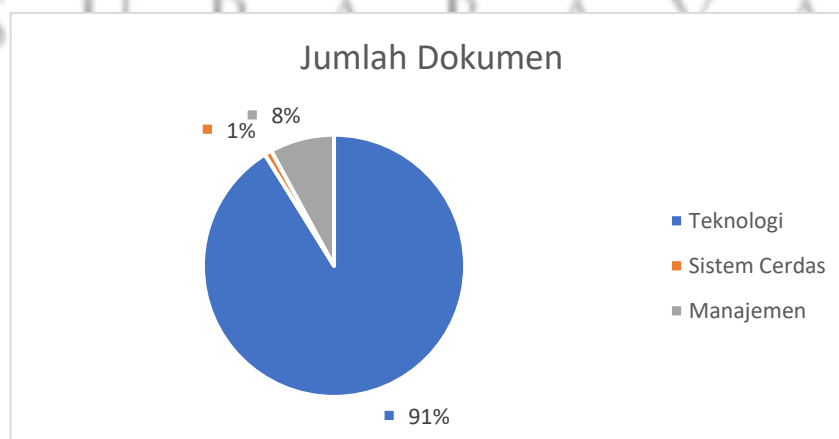
4.2.6 Analisis Topik pada tahun 2020

Pada hasil proses LDA tahun 2020 juga dilakukan interpretasi untuk mengetahui topik-topik sistem informasi apa yang telah dihasilkan. Tabel 4.22 Merupakan hasil interpretasi topik-topik pada tahun 2020.

Tabel 4. 22 Hasil interpretasi topik tahun 2020

Topic id	Kata Penyusun (Top 10 kata)	Hasil Interpretasi	
0	"data", "method", "process", "study", "test", "algorithm", "learn", "application", "determine", "decision"	Manajemen	53
1	"application", "method", "study", "process", "service", "user", "design", "data", "analysis", "technology"	Sistem Cerdas	6
2	"data", "process", "study", "test", "method", "model", "analysis", "technology", "development", "design"	Teknologi	608

Dengan mengacu pada Tabel 4.22 yang menampilkan jumlah dokumen terkait ke tiap topik, maka persentase jumlah dokumen terhadap topik di 2020 ada pada Gambar 4.21 berikut.



Gambar 4. 21 Persentase jumlah dokumen terhadap topik di 2020

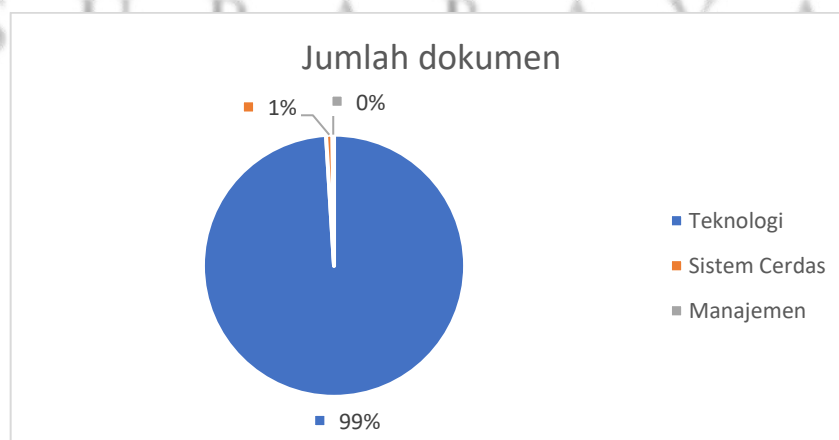
4.2.7 Analisis Topik pada tahun 2021

Pada hasil proses LDA tahun 2021 dilakukan interpretasi topik guna membantu analisis tren topik sistem informasi di proses selanjutnya. Tabel 4.23 Merupakan hasil interpretasi topik-topik pada tahun 2021.

Tabel 4. 23 Hasil interpretasi topik tahun 2021

<i>Topic id</i>	<i>Kata Penyusun (Top 10 kata)</i>	<i>Hasil Interpretasi</i>	<i>Jumlah Dokumen</i>
0	"study", "method", "application", "learn", "process", "data", "test", "management", "model", "risk"	Manajemen	59
1	"data", "study", "method", "process", "algorithm", "test", "accuracy", "model", "time", "application"	Sistem Cerdas	6
2	"process", "service", "technology", "method", "design", "application", "study", "development", "data", "model"	Teknologi	752

Dengan mengacu pada Tabel 4.23 yang menampilkan jumlah dokumen terkait ke tiap topik, maka persentase jumlah dokumen terhadap topik di 2015 ada pada Gambar 4.22 berikut.



Gambar 4. 22 Persentase jumlah dokumen terhadap topik di 2021

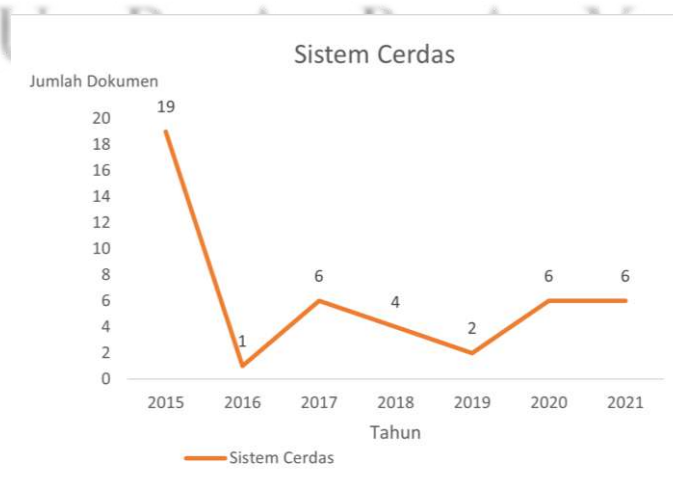
4.3 Analisis Tren Tahunan

Dengan menggunakan hasil interpretasi dari topik-topik yang dihasilkan oleh proses LDA, penelitian dilanjutkan dengan mencari tren tahunan dari topik-topik tersebut. Untuk mendapatkan tren naik dan turun terhadap tiap-tiap topik, penelitian ini mengambil nilai jumlah dokumen terhadap topik di tiap tahunnya untuk dijadikan acuan proses analisis tren. Fungsi *get_document_topics()* akan menghasilkan probabilitas topik dari setiap dokumen di dalam *corpus*. Dari tiap topik di *corpus* yang kemudian akan dihitung jumlah dokumen yang terkait ke dalam topik untuk mendapatkan peringkat topik apa yang paling dominan di dalam *corpus* tersebut. Tabel 4.24 merupakan pemetaan jumlah dokumen terkait ke tiap *topic* di dalam *corpus*.

Tabel 4. 24 Jumlah dokumen terkait ke tiap topik di dalam corpus

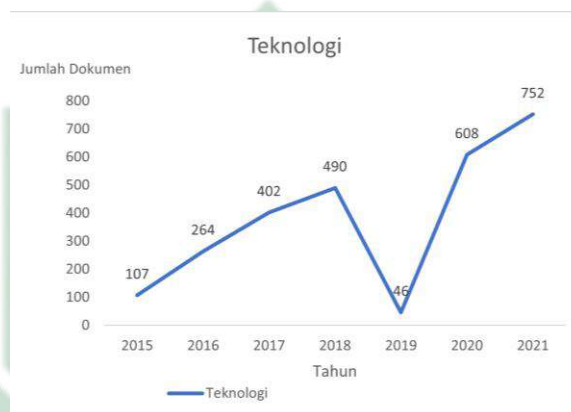
<i>Corpus</i>	<i>Teknologi</i>	<i>Sistem Cerdas</i>	<i>Manajemen</i>
2015	107	19	55
2016	264	1	22
2017	402	6	41
2018	490	4	42
2019	46	2	545
2020	608	6	53
2021	752	6	59

Dari Tabel 4.24 di atas, dapat terlihat bahwa dari ketiga topik tersebut terdapat kenaikan dan penurunan dalam tren pertahunnya. Gambar 4.23 merupakan tren tahunan untuk topik Sistem Cerdas.



Gambar 4. 23 Tren tahunan topik Sistem Cerdas

Berdasarkan Gambar 4.23 dapat dilihat bahwa penelitian yang meneliti tentang topik sistem cerdas sudah tergolong rendah/sedikit. Topik sistem cerdas mengalami penurunan tren jumlah penelitian di tahun 2016. Kemudian di tahun-tahun selanjutnya, topik sistem cerdas mengalami tren naik-turun yang tidak signifikan dan juga tidak konsisten sehingga tidak banyak perubahan pada angka penelitian di tiap tahunnya. Berikutnya, Gambar 4.24 merupakan tren tahunan untuk topik Teknologi.



Gambar 4. 24 Tren tahunan topik Teknologi

Berdasarkan Gambar 4.24 dapat terlihat bahwa topik teknologi mengalami kenaikan jumlah dokumen penelitian di tahun 2016 hingga 2018. Namun pada tahun 2019, topik teknologi mengalami penurunan jumlah penelitian yang cukup signifikan sebelum akhirnya jumlah penelitian topik teknologi kembali mengalami kenaikan di tahun 2020 dan berlanjut di tahun 2021. Terakhir, Gambar 4.25 merupakan tren tahunan untuk topik Manajemen.



Gambar 4. 25 Tren tahunan topik Manajemen

Berdasarkan Gambar 4.25 Dapat terlihat bahwa topik Manajemen mengawali tren di 2015 dengan jumlah yang cukup rendah namun jumlah tersebut cukup stabil hingga tahun 2018. Pada tahun 2019, berbanding terbalik dengan topik Teknologi yang mengalami penurunan dalam jumlah dokumen penelitian, topik Manajemen justru mengalami peningkatan yang signifikan hingga menyentuh angka 545 jumlah dokumen penelitian. Namun kemudian topik Manajemen mengalami penurunan yang signifikan juga di tahun 2020 yang berlanjut hingga tahun selanjutnya. Dibandingkan dengan 2 topik lainnya, topik Sistem Cerdas merupakan topik yang masih sedikit penelitiannya. Sedangkan topik Teknologi dapat dikatakan sebagai topik yang paling populer untuk diangkat sebagai bahan penelitian sistem informasi dengan melihat jumlah penelitian yang cukup tinggi di tiap tahunnya. Untuk topik Manajemen dapat dikatakan masih tergolong kurang populer karena penelitian yang meneliti topik Manajemen jumlahnya masih tidak cukup banyak dibandingkan dengan topik Teknologi. Dapat disimpulkan bahwa topik Teknologi merupakan topik yang umum diangkat sebagai bahan penelitian di dalam jurnal-jurnal sistem informasi. Sedangkan topik Sistem Cerdas dan topik Manajemen merupakan 2 topik yang masih belum banyak diangkat sebagai bahan penelitian di dalam jurnal tiap tahunnya, sehingga hal tersebut membuka kesempatan para peneliti di masa depan yang ingin melakukan penelitian di bidang sistem informasi.

UIN SUNAN AMPEL
S U R A B A Y A

BAB V PENUTUP

5.1 Kesimpulan

Kesimpulan pada penelitian ini yaitu *topic coherence* digunakan untuk membantu menemukan jumlah topik yang akan digunakan oleh model LDA. Hasil dari *topic coherence* menunjukkan jumlah topik 6 adalah jumlah topik dengan nilai *coherence* terbesar. Namun setelah diterapkan dan divisualisasikan ke dalam model LDA, jumlah topik 6 masih menghasilkan beberapa topik yang saling beririsan satu sama lain. Oleh karena itu dilakukan pengurangan jumlah topik hingga mendapat jumlah topik 3 adalah jumlah topik yang optimal untuk model LDA. Konsultasi pada ahli bidang sistem informasi menghasilkan 3 topik dalam bidang sistem informasi yang bisa digunakan pada model LDA penelitian ini yaitu topik Teknologi, topik Sistem Cerdas, dan topik Manajemen).

Pada analisis tren tahunan didapatkan bahwa dari ketiga topik yang ada, topik Teknologi merupakan topik yang umum diangkat sebagai bahan penelitian di bidang sistem informasi. Sedangkan topik Sistem Cerdas dan topik Manajemen merupakan dua topik yang masih kurang populer dibandingkan dengan topik Teknologi dengan mengacu pada jumlah penelitiannya. Oleh karena itu, hal tersebut bisa menjadi kesempatan untuk peneliti yang melakukan penelitian dalam bidang sistem informasi di masa depan untuk mengangkat topik-topik yang masih kurang populer.

5.2 Saran Pengembangan

Untuk saran pengembangan penelitian ini, yang pertama adalah menyempurnakan proses *pre-processing* yang ada. Karena masih terdapat beberapa proses *pre-processing* yang bisa dilakukan untuk hasil yang lebih optimal seperti pembentukan *bigram*, *trigram*, dan *n-gram*. Yang kedua, melakukan metode lain untuk menentukan jumlah topik yang akan digunakan dalam penelitian, misalnya penggunaan metode *topic perplexity* atau menggunakan *topic coherence* dengan metode *measurement* yang lain contohnya Umass, npmi, uci, dll. untuk mendapatkan jumlah topik yang lebih pas dan topik yang lebih baik.

DAFTAR PUSTAKA

- Aditya, B. R. (2015). Penggunaan Web Crawler Untuk Menghimpun Tweets dengan Metode Pre-Processing Text Mining. *JURNAL INFOTEL - Informatika Telekomunikasi Elektronika*, 7(2), 93. <https://doi.org/10.20895/infotel.v7i2.35>
- Afuan, L. (2013). Stemming Dokumen Teks Bahasa Indonesia Menggunakan Algoritma Porter. *Telematika*, 6(2), 34–40.
- Alfanzar, A. I. (2019). *Topic modelling skripsi menggunakan metode latent dirichlet allocation*.
- Aliyanto, D. (2017). Pemingkatan Perguruan Tinggi Menggunakan Metode Probabilistic Latent Semantic Analysis (pLSA) untuk Mengukur Tingkat Kesiapterapan Teknologi di Indonesia. *Jurnal Teknik ITS*, 6(2). <https://doi.org/10.12962/j23373539.v6i2.24221>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3(4–5), 993–1022. <https://doi.org/10.1016/b978-0-12-411519-4.00006-9>
- Catford, J. C. (1965). A Linguistic Theory of Translation. In *Pliego de Yuste*. <http://www.pliegosdeyuste.eu/n4pliegos/eugeneanida.pdf>
- Deepu, S., Pethuru, R., & Rajaraajeswari, S. (2016). A Framework for Text Analytics using the Bag of Words (BoW) Model for Prediction. *International Conference on Innovations in Computing & Networking*, 320–323.
- Fadli, A. (2003). Konsep Data Mining. *Konsep Data Mining*, 1–9.
- Gerintya, S. (2017). *Kondisi Dunia Penelitian di Indonesia*. <https://tirto.id/kondisi-dunia-penelitian-di-indonesia-cvvyj>
- Jeyaraj, A., & Zadeh, A. H. (2019). Evolution of information systems research: Insights from topic modeling. *Information and Management*, 103207. <https://doi.org/10.1016/j.im.2019.103207>
- Moertini, V. S. (2002). *DATA MINING SEBAGAI SOLUSI BISNIS*. 7(1), 44–56.
- Mujilawati, S. (2016). Pre-Processing Text Mining Pada Data Twitter. *Seminar Nasional Teknologi Informasi Dan Komunikasi, 2016*(Sentika), 2089–9815.
- Nastiti, K. R. (2019). *PEMODELAN TOPIK UNTUK PENELITIAN DI BIDANG INFORMATIKA MENGGUNAKAN METODE LATENT DIRICHLET ALLOCATION*.
- Putra, I. M. K. B. (2017). Analisis Topik Informasi Publik Media Sosial di Surabaya Menggunakan Pemodelan Latent Dirichlet Allocation (LDA). *Jurnal Teknik ITS*. <https://doi.org/10.12962/j23373539.v6i2.23205>

- Putranti, N. D., & Winarko, E. (2014). Analisis Sentimen Twitter untuk Teks Berbahasa Indonesia dengan Maximum Entropy dan Support Vector Machine. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 8(1), 91–100. <https://doi.org/10.22146/ijccs.3499>
- Putri, W. T. H., & Hendrowati, R. (2018). Penggalan Teks Dengan Model Bag of Words Terhadap. *Jurnal Muara Sains, Teknologi, Kedokteran, Dann Ilmu Kesehatan*, 2(1), 129–138.
- Rahm, E., & Do, H. H. (2000). *Data Cleaning: Problems and Current Approaches I*. <http://list.research.microsoft.com/scripts/lyris.pl?enter=debull>.
- Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the space of topic coherence measures. *WSDM 2015 - Proceedings of the 8th ACM International Conference on Web Search and Data Mining*, 399–408. <https://doi.org/10.1145/2684822.2685324>
- Rozaliya, B. Z. H. (2018). ANALISIS TOPIK PADA SUARA KONSUMEN PT ANGKASA PURA 1 (PERSERO) CABANG JUANDA DENGAN MENGGUNAKAN PEMODELAN GAUSSIAN LATENT DIRICHLET ALLOCATION (GLDA).
- Sasmita, R. A. (2014). PEMANFAATAN ALGORITMA TF / IDF UNTUK SISTEM INFORMASI e-COMPLAINT HANDLING.
- Triawati, C. (2009). *Metode Pembobotan Statistical Concept Based Untuk Klustering Dan Kategorisasi*.
- Utami, K. P. (2017). Analisis topik data media sosial twitter menggunakan model topik latent dirichlet allocation.
- Utomo, M. S. (2015). Stopword Dinamis dengan Pendekatan Statistik. *Jurnal Informatika Upgris (JIU)*, 140–148.
- Zulhanif. (2016). Pemodelan Topik Dengan Latent Dirichlet Allocation. *Seminar Nasional Pendidikan Matematika*, 1–8.