



TUGAS AKHIR - KS141501

**ANALISIS TOPIK DAN *AUTHOR* PUBLIKASI PADA
REPOSITORI *OPEN ACCESS JOURNAL OF INFORMATION
SYSTEMS* DENGAN MENGGUNAKAN METODE *AUTHOR-
TOPIC MODELS***

***TOPICS AND AUTHOR PUBLICATION ANALYSIS ON OPEN
ACCESS JOURNAL OF INFORMATION SYSTEMS
REPOSITORY USING AUTHOR-TOPIC MODELS***

**SITI ORYZA KHAIRUNNISA
NRP 5213 100 110**

**Dosen Pembimbing:
Renny Pradina Kusumawardani, S.T., M.T., SCJP**

**DEPARTEMEN SISTEM INFORMASI
Fakultas Teknologi Informasi dan Komunikasi
Institut Teknologi Sepuluh Nopember
Surabaya 2018**



ITS
Institut
Teknologi
Sepuluh Nopember

TUGAS AKHIR - KS141501

**ANALISIS TOPIK DAN *AUTHOR* PUBLIKASI PADA
REPOSITORI *OPEN ACCESS JOURNAL OF INFORMATION
SYSTEMS* DENGAN MENGGUNAKAN METODE *AUTHOR-
TOPIC MODELS***

**SITI ORYZA KHAIRUNNISA
NRP 5213 100 110**

**Dosen Pembimbing:
Renny Pradina Kusumawardani, S.T., M.T., SCJP**

**DEPARTEMEN SISTEM INFORMASI
Fakultas Teknologi Informasi dan Komuniiasi
Institut Teknologi Sepuluh Nopember
Surabaya 2018**



FINAL PROJECT - KS 141501

***TOPICS AND AUTHOR PUBLICATION ANALYSIS ON OPEN
ACCESS JOURNAL OF INFORMATION SYSTEMS
REPOSITORY USING AUTHOR-TOPIC MODELS***

**SITI ORYZA KHAIRUNNISA
NRP 5213 100 110**

Supervisor:

Renny Pradina Kusumawardani, S.T., M.T., SCJP

DEPARTMENT OF INFORMATION SYSTEMS

Faculty of Information Technology and Communication

Institut Teknologi Sepuluh Nopember

Surabaya 2018

LEMBAR PENGESAHAN

ANALISIS TOPIK DAN *AUTHOR* PUBLIKASI PADA REPOSITORI *OPEN ACCESS JOURNAL OF INFORMATION SYSTEMS* DENGAN MENGUNAKAN METODE *AUTHOR-TOPIC MODELS*

TUGAS AKHIR

Disusun untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer
pada

Departemen Sistem Informasi
Fakultas Teknologi Informasi dan Komunikasi
Institut Teknologi Sepuluh Nopember

Oleh:

SITI ORYZA KHAIRUNNISA

5213 100 110

Surabaya, 17 Januari 2018

**PLH KEPALA
DEPARTEMEN SISTEM INFORMASI**

Edwin Riksakomara, S.Kom., M.T.

NIP 196907252003121001



LEMBAR PERSETUJUAN

ANALISIS TOPIK DAN *AUTHOR* PUBLIKASI PADA REPOSITORI *OPEN ACCESS JOURNAL OF INFORMATION SYSTEMS* DENGAN MENGUNAKAN METODE *AUTHOR-TOPIC MODELS*

TUGAS AKHIR

Disusun Untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer
pada

Departemen Sistem Informasi

Fakultas Teknologi Informasi dan Komunikasi
Institut Teknologi Sepuluh Nopember


Oleh :

SITI ORYZA KHAIRUNNISA


5213 100 110

Disetujui Tim Penguji: Tanggal Ujian : 12 Januari 2018
Periode Wisuda : Maret 2018


Renny Pradina Kusumawardani, S.T., M.T., SCJP


(Pembimbing I)

Nur Aini Rakhmawati, S.Kom., M.Sc.Eng., Ph.D


(Penguji I)

Irmasari Hafidz, S.Kom., M.Sc


(Penguji II)

**ANALISIS TOPIK DAN *AUTHOR* PUBLIKASI
PADA REPOSITORY *OPEN ACCESS JOURNAL OF
INFORMATION SYSTEMS* DENGAN
MENGUNAKAN METODE *AUTHOR-TOPIC
MODELS***

Nama Mahasiswa : Siti Oryza Khairunnisa
NRP : 5213 100 110
Departemen : Sistem Informasi
Pembimbing 1 : Renny Pradina K., S.T.,
M.T., SCJP

ABSTRAK

Open Access Journal of Information Systems (OAJIS) merupakan repositori terbuka yang disediakan oleh Jurusan Sistem Informasi ITS untuk seluruh peneliti bidang sistem informasi di seluruh dunia. Dalam menentukan reviewer yang sesuai untuk mengevaluasi publikasi yang masuk, masih dilakukan secara manual berdasarkan publikasi reviewer dan disesuaikan dengan penelitian yang akan dievaluasi. Dengan sistem manual seperti ini, penentuan reviewer terkadang masih kurang tepat sehingga berdampak pada hasil evaluasi penelitian yang masuk ke OAJIS.

Author-Topic Models merupakan sebuah metode pengembangan dari LDA, dengan tambahan integrasi informasi mengenai author dari suatu dokumen pada topic modeling. Metode ini dapat digunakan untuk mengetahui tren topik terkini dari waktu ke waktu, mengetahui topik dokumen berdasarkan author, dan penentuan topik dan author untuk dokumen baru yang tidak terdapat dalam kumpulan dokumen yang dimiliki.

Dengan kondisi pada OAJIS, dilakukan analisa topik dari setiap penelitian yang masuk menggunakan metode *Author-*

Topic Models sehingga dapat memudahkan pengelola dalam menentukan reviewer.

Berdasarkan eksperimen pemodelan topik yang telah dilakukan, diketahui bahwa data yang telah skenario stemming menghasilkan model yang paling baik. Dengan korelasinya pada kebutuhan dalam menentukan reviewer, dilakukan pengujian menggunakan similaritas vektor probabilitas. Dari hasil pengujian diketahui bahwa model skenario stemming dan jumlah topik 50 dapat menghasilkan kemampuan prediksi yang lebih akurat, dengan nilai perplexity 128.71 dan nilai topic coherence -1.528.

Kata Kunci: Author-Topic Models, OAJIS, Topic Modeling, Probabilistic Vector Similarity.

***TOPICS AND AUTHOR PUBLICATION ANALYSIS
ON OPEN ACCESS JOURNAL OF INFORMATION
SYSTEMS REPOSITORY USING AUTHOR-TOPIC
MODELS***

Nama Mahasiswa : **Siti Oryza Khairunnisa**
NRP : **5213 100 110**
Departemen : **Sistem Informasi**
Pembimbing 1 : **Renny Pradina K., S.T.,
M.T., SCJP**

ABSTRACT

Open Access Journal of Information Systems (OAJIS) is an open-repository which is provide by Information Systems Department of Institut Teknologi Sepuluh Nopember for all researchers in information systems field around the world. In considering which reviewer is suitable for evaluating when a new publication is registered, the process is still done manually based on reviwer's publication compared to the new one. A manual process like this makes reviewer assigning process less precise so it will impact the result of evaluation.

Author-Topic Models is a developed method from LDA, with additional information integration about the authors of the document in topic modeling. This method can be used to identify the trend of recent topics from time to time, knowing a document's topic based on its authors, and predicting topics and authors for unknown documents.

With the condition of OAJIS, we conducted a topic and author analysis for the publication in the repository using Author-Topic Models method so it will make the reviewer assigning process easier.

Based on the topic modeling experiment, it is known that data with stemming scenario resulted the best model. With its correlation to reviewer assigning process, a testing was

conducted using probability vector similarity. From the testing result, model with stemming scenario and 50 number of topics could result better prediction capability for the reviewer, with the average perplexity score 128.71 and -1.528 topic coherence score.

Keyword: Author-Topic Models, OAJIS, Topic Modeling, Probabilistic Vector Similarity.

KATA PENGANTAR

Puji dan syukur penulis tuturkan ke hadirat Allah SWT, Tuhan Semesta Alam yang telah memberikan kekuatan dan hidayah-Nya kepada penulis sehingga penulis mendapatkan kelancaran menyelesaikan tugas akhir yang merupakan salah satu syarat kelulusan pada Departemen Sistem Informasi, Fakultas Teknologi Informasi dan Komunikasi, Institut Teknologi Sepuluh Nopember Surabaya.

Terima kasih penulis sampaikan kepada pihak-pihak yang telah mendukung, memberikan saran, motivasi, semangat, dan bantuan baik berupa materiil maupun moril demi tercapainya tujuan pembuatan tugas akhir ini. Tugas akhir ini tidak akan pernah terwujud tanpa bantuan dan dukungan dari berbagai pihak yang sudah meluangkan waktu, tenaga dan pikirannya. Secara khusus penulis akan menyampaikan ucapan terima kasih yang sebanyak-banyaknya kepada:

1. Bapak Moh Aminullah dan Ibu Budi Irmawati selaku kedua orang tua serta Muhammad Radite Rahmatullah selaku adik kandung dari penulis yang tiada henti memberikan dukungan dan semangat secara lahir dan batin.
2. Ibu Renny Pradina Kusumawardani., S.T., M.T., SCJP, selaku dosen pembimbing dan sebagai narasumber yang senantiasa meluangkan waktu dan tenaga, memberikan ilmu dan petunjuk yang sangat berarti bagi penulis, serta memotivasi untuk kelancaran pengerjaan tugas akhir.
3. Ibu Mahendrawathi ER., S.T., M.Sc., Ph.D., selaku dosen wali yang selalu membantu penulis dalam masa perkuliahan sejak mahasiswa baru hingga tugas akhir.
4. Ibu Nur Aini R., S.Kom., M.Sc.Eng., Ph.D., dan Ibu Irmasari Hafidz, S.Kom., M.Sc., selaku dosen penguji yang telah memberikan saran dan kritik untuk perbaikan tugas akhir.
5. Seluruh dosen Departemen Sistem Informasi ITS yang telah memberikan ilmu yang bermanfaat kepada penulis.

6. Kirana Gita Larasati, Ashma Hanifah Shalihah, Safrina Kharisma Imandani, dan Daniel Surya Anjas Marbun selaku sahabat terdekat yang telah memberikan dukungan yang sangat berarti bagi penulis.
7. Fitriyani Aditiya, Bianca Adriennawati, dan Unsa Sabrina yang telah mendukung dan menemani penulis sejak masih di Sekolah Dasar hingga tugas akhir ini dapat diselesaikan.
8. Stezar Priansya, Achmad Saiful, I Made Kusnanta, Tetha Valianta, Hanum Fitriani, dan Ari Agustina yang memberikan pencerahan secara praktek dan teori terkait pengerjaan tugas akhir.
9. Opor, Fia, Alden, Septi, Paty, Rara, Nita, Zuli, Pras, Putra, Fachrur, dan Fata yang mendukung, menemani, dan berjuang bersama penulis dalam pengerjaan tugas akhir.
10. Tatan, Fikry, Faiz, Guntur, Dije, Gradi, Satria, Fauzul, dan Arep yang senantiasa menghibur dan mendukung penulis dalam pengerjaan tugas akhir.
11. Rekan-rekan BEM FTIf, Pagelaran Seni ITS EXPO, dan Kementrian Sosial Masyarakat BEM ITS yang telah memberikan banyak kenangan dan pembelajaran yang sangat berharga semasa kuliah.
12. Pinka, Azhar, Ipeh, Izzat, Yoga, Faathir, Stephanie, Sekar, Rici, Hilmi, Adam, Kinan, Alwan, Kak Astin, Kak Vani, Kak Raras, Kak Fajar, Kak Indah, Kak Madam, Kak Itsna, Mbak Athifah, Mas Angga, Bang Syahril, Mbak Amel, dan Mbak Fida yang menemani dan selalu membantu penulis dengan ikhlas semasa menjalani kehidupan sebagai mahasiswa program *exchange* di Tohoku University.

Penyusunan laporan ini masih jauh dari kata sempurna sehingga penulis menerima kritik maupun saran yang membangun untuk perbaikan di masa yang akan datang. Semoga buku tugas akhir ini dapat memberikan manfaat bagi pembaca.

Surabaya, 7 Januari 2018
Penulis,

Siti Oryza Khairunnisa

DAFTAR ISI

LEMBAR PENGESAHAN.....	iv
LEMBAR PERSETUJUAN.....	v
ABSTRAK.....	vi
ABSTRACT.....	viii
KATA PENGANTAR.....	x
DAFTAR ISI.....	xii
DAFTAR GAMBAR.....	xvi
DAFTAR KODE.....	xvii
DAFTAR TABEL.....	xviii
DAFTAR DIAGRAM.....	xix
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Perumusan Masalah.....	2
1.3 Batasan Masalah.....	2
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
1.6 Relevansi.....	3
BAB II TINJAUAN PUSTAKA.....	5
2.1 Studi Sebelumnya.....	5
2.2 Dasar Teori.....	9
2.2.1 <i>Natural Language Processing</i>	9
2.2.2 <i>Topic Modeling</i>	11
2.2.3 <i>Latent Dirichlet Allocation (LDA)</i>	11
2.2.4 <i>Author-Topic Models</i>	12
2.2.5 <i>Part-of-Speech Tag</i>	14
2.2.6 Evaluasi Model Berdasarkan Nilai <i>Perplexity</i> ...	15
2.2.7 Evaluasi Model Berdasarkan <i>Topic Coherence</i> .	15
2.2.8 Perhitungan Nilai <i>Similarity Matrix</i> berdasarkan <i>Hellinger Distance</i>	16
2.2.9 <i>Open Access Journal of Information Systems</i> (<i>OAJIS</i>).....	16
BAB III METODOLOGI.....	17
3.1 Tahapan Pelaksanaan Tugas Akhir.....	17

3.1.1	Identifikasi Masalah.....	18
3.1.2	Studi Literatur	18
3.1.3	Proses Pengumpulan Data.....	19
3.1.4	Part-of-Speech Tag	21
3.1.5	Pra-proses Data	21
3.1.6	<i>Topic Modeling</i> dengan <i>Author-Topic Model</i> ...	25
3.1.7	Evaluasi Model	26
3.1.8	Analisis Hasil.....	26
3.1.9	Dokumentasi	26
BAB IV PERANCANGAN		27
4.1	Pengambilan Data	27
4.2	Metodologi Implementasi Penelitian.....	27
4.2.1	Mempersiapkan Data	28
4.2.2	<i>Part-of-speech tag</i>	30
4.2.3	Pra-Proses Data.....	31
4.2.4	<i>Topic Modeling</i>	32
4.2.5	Validasi Model.....	33
4.3	Perancangan Pengujian Model	33
4.3.1	Probabilitas Kata dalam Dokumen per Topik ...	34
4.3.2	Perhitungan Kemiripan Dokumen dan <i>Author</i> berdasarkan Similaritas Vektor Probabilitas menggunakan <i>Hellinger Distance</i>	35
BAB V IMPLEMENTASI.....		37
5.1	Lingkungan Implementasi.....	37
5.2	Mempersiapkan Data	38
5.2.1	Memberikan <i>tag</i> pada data untuk skenario data <i>topic modeling</i> dengan <i>noun part-of-speech tag</i>	38
5.2.2	Memuat Data	41
5.2.3	Membersihkan Data	43
5.3	Pra-proses Data	44
5.3.1	Pendefinisian <i>stopword</i>	44
5.3.2	<i>Case folding</i> , <i>stemming</i> , tokenisasi, dan penghapusan <i>stopword</i>	45
5.3.3	Pembuatan frasa.....	46
5.3.4	Pembuatan <i>dictionary</i> dari dokumen.....	47

5.3.5	Pembuatan dokumen matriks atau <i>corpus</i>	47
5.4	Pemodelan Topik dengan metode <i>Author-Topic Models</i>	48
5.4.1	Pembentukan model topik	48
5.4.2	Eksperimen pemodelan topik dengan <i>Author-Topic Model</i>	48
5.5	Validasi Model Topik	51
5.5.1	Memanggil hasil <i>Top Topics</i>	51
5.5.2	Rata-rata <i>Coherence Score</i>	52
5.6	Pengujian model dengan menggunakan nilai similaritas vektor probabilitas.....	52
5.6.1	Memuat data pengujian	53
5.6.2	Perhitungan rata-rata probabilitas kata dalam dokumen untuk setiap topik	53
5.6.3	Menghitung nilai similaritas vektor probabilitas	
	55	
BAB VI	HASIL DAN PEMBAHASAN	57
6.1	Mempersiapkan Data	57
6.1.1	Memuat Data.....	57
6.1.2	Pemberian <i>tag</i> pada data dengan POS Tag Indonesia untuk skenario data <i>topic modeling</i> dengan <i>noun part-of-speech tag</i>	57
6.2	Pra-Proses Data	58
6.3	Pembuatan <i>Dictionary</i> dari Dokumen	58
6.4	Pemodelan dengan <i>Author-Topic Model</i>	59
6.4.1	Penentuan Jumlah <i>Passes</i>	59
6.4.2	Penentuan Jumlah Topik	60
6.5	Validasi Model Topik	65
6.5.1	<i>Top Topics</i>	65
6.5.2	Nilai <i>Topic Coherence</i>	65
6.6	Pengujian Model	68
6.6.1	Analisis Hasil berdasarkan Skenario Data.....	71
6.6.2	Analisis Hasil berdasarkan <i>Minimal Author Size</i>	

6.6.3 Validasi <i>Keyword</i> dengan Distribusi kata dalam Topik.....	80
BAB VII KESIMPULAN DAN SARAN.....	83
7.1 Kesimpulan.....	83
7.2 Saran dan Penelitian Selanjutnya	84
DAFTAR PUSTAKA	85
BIODATA PENULIS	87
LAMPIRAN A.....	A-1
LAMPIRAN B	B-1
LAMPIRAN C	C-1
LAMPIRAN D.....	D-1

DAFTAR GAMBAR

Gambar 2.2.1 Model LDA.....	12
Gambar 2.2.2 Model <i>Author-Topic Models</i>	13
Gambar 2.2.3 Contoh hasil <i>tagging</i> menggunakan POS Tag Indonesia.....	14
Gambar 3.1.1 Contoh <i>file paper</i> dalam bentuk <i>.pdf</i>	20
Gambar 3.1.2 Contoh data <i>file paper</i> yang telah dikonversi menjadi <i>.txt</i>	20
Gambar 3.1.3 Hasil <i>tagging</i> menggunakan POS Tag Indonesia.....	21
Gambar 3.1.4 Sampel hasil <i>topic modeling</i>	25
Gambar 4.2.1 <i>list author</i>	29
Gambar 5.2.1 Susunan direktori penggunaan POS Tag Indonesia.....	39
Gambar 5.2.2 Hasil proses <i>tagging</i> dengan POS Tag Indonesia.....	40
Gambar 6.6.1 Contoh hasil pengujian similaritas vektor probabilitas.....	69
Gambar 6.6.2 Hasil pengujian dokumen 3.....	74
Gambar 6.6.3 Hasil pengujian dokumen 1.....	76

DAFTAR KODE

Kode 5.1 Menjalankan POS Tag Indonesia.....	40
Kode 5.2 Menghapus kata dengan <i>tag</i> selain 'NN' dan 'NNP'.....	41
Kode 5.3 <i>Import library logging, os, re, csv</i>	41
Kode 5.4 Memuat <i>list</i> dokumen ke dalam <i>Python</i>	42
Kode 5.5 Memuat <i>list author</i> ke dalam <i>Python</i>	43
Kode 5.6 Membuat hubungan <i>author2doc</i> antara <i>list author</i> dengan <i>list</i> dokumen	43
Kode 5.7 Membersihkan data.....	44
Kode 5.8 Pendefinisian <i>stopword list</i>	45
Kode 5.9 Pembuatan <i>Stemmer</i> dan <i>Tokenizer</i>	45
Kode 5.10 <i>Case Folding, Stemming, Tokenisasi, dan penghapusan stopwords dari dokumen</i>	46
Kode 5.11 Membuat frasa kata.....	47
Kode 5.12 Membuat <i>dictionary unique tokens</i> dari dokumen.....	47
Kode 5.13 Pembuatan <i>corpus</i> berdasarkan <i>dictionary</i> dan dokumen.....	48
Kode 5.14 Pembentukan topik model dengan <i>Author-Topic Model</i>	48
Kode 5.15 Penentuan Jumlah <i>Passes</i>	50
Kode 5.16 Penentuan Jumlah Topik.....	50
Kode 5.17 Menyimpan Model.....	51
Kode 5.18 <i>Load</i> model.....	51
Kode 5.19 Memanggil hasil <i>Top Topics</i>	52
Kode 5.20 Menghitung rata-rata <i>coherence score</i>	52
Kode 5.21 Memuat data pengujian.....	53
Kode 5.22 Perhitungan rata-rata probabilitas kata per topik.....	54
Kode 5.23 Mengurutkan dan mengambil nilai probabilitas 5 teratas	54
Kode 5.24 Perhitungan nilai similaritas vektor probabilitas menggunakan <i>Hellinger Distance</i>	55
Kode 5.25 Membuat tabel nilai similaritas vektor probabilitas antara dokumen dan <i>author</i>	56
Kode 5.26 Memanggil distribusi topik <i>author</i>	56

DAFTAR TABEL

Tabel 2.1.1 Studi sebelumnya.....	5
Tabel 3.1.1 Contoh data dari repositori <i>OAJIS</i>	19
Tabel 3.1.2 <i>Data cleaning</i>	22
Tabel 3.1.3 <i>Case folding</i>	23
Tabel 3.1.4 <i>Stemming</i>	23
Tabel 3.1.5 <i>Stopwords removal</i>	24
Tabel 3.1.6 <i>Tokenization</i>	25
Tabel 5.1.1 Spesifikasi Komputer	37
Tabel 5.1.2 Teknologi yang digunakan untuk mengembangkan model.....	38
Tabel 6.1.1 Jumlah data pelatihan	57
Tabel 6.1.2 Jumlah data <i>tag</i>	57
Tabel 6.2.1 Perubahan jumlah kata setelah pra-proses	58
Tabel 6.3.1 Jumlah <i>unique tokens</i>	59
Tabel 6.4.1 Rata-rata dan standar deviasi nilai perplexity model	65
Tabel 6.6.1 <i>Recall author</i> dokumen pengujian	70
Tabel 6.6.2 Distribusi topik dokumen 3	75
Tabel 6.6.3 Distribusi topik <i>author</i> Hanim Maria Astuti	75
Tabel 6.6.4 Distribusi topik <i>author</i> Eko Nugroho	75
Tabel 6.6.5 Distribusi topik dokumen 1	77
Tabel 6.6.6 Distribusi topik <i>author</i> Febriliyan Samopa.....	77
Tabel 6.6.7 Distribusi Topik <i>author</i> Annis Paramita Dilla	77
Tabel 6.6.8 Dokumen 1	78
Tabel 6.6.9 Dokumen <i>author</i> Annis Paramita Dilla	79

DAFTAR DIAGRAM

Diagram 3.1.1 Tahapan pelaksanaan tugas akhir.....	17
Diagram 4.2.1 Alur mempersiapkan data	28
Diagram 4.2.2 Alur tahap pemberian tag pada data.....	30
Diagram 4.2.3 Alur tahap pra-proses data	31
Diagram 4.2.4 Alur pembentukan <i>dictionary</i> dan <i>corpus</i>	32
Diagram 4.2.5 Alur pementukkan model	33
Diagram 4.3.1 Alur perhitungan vektor probabilitas dokumen	34
Diagram 6.4.1 Analisa nilai <i>perplexity</i> untuk menentukan jumlah <i>passes</i>	60
Diagram 6.4.2 Rata-rata nilai <i>perplexity</i> dalam 30 kali percobaan	61
Diagram 6.4.3 Standar deviasi nilai <i>perplexity</i> dalam 30 kali percobaan	61
Diagram 6.4.4 Rata-rata nilai <i>perplexity</i> dalam 30 kali percobaan	62
Diagram 6.4.5 Standar deviasi nilai <i>perplexity</i> dalam 30 kali percobaan	63
Diagram 6.4.6 Rata-rata nilai <i>perplexity</i> dalam 30 kali percobaan	63
Diagram 6.4.7 Standar deviasi nilai <i>perplexity</i> dalam 30 kali percobaan	64
Diagram 6.5.1 Rata-rata nilai <i>topic coherence</i>	66
Diagram 6.5.2 Standar deviasi nilai <i>topic coherence</i>	66
Diagram 6.5.3 Nilai <i>topic coherence</i> tertinggi.....	67
Diagram 6.5.4 Nilai <i>topic coherence</i> terendah	67
Diagram 6.6.1 Histogram jumlah dokumen yang ditulis oleh <i>author</i>	72

BAB I

PENDAHULUAN

Bab pendahuluan ini menguraikan proses identifikasi masalah penelitian yang meliputi latar belakang masalah, perumusan masalah, batasan masalah, tujuan tugas akhir, manfaat kegiatan tugas akhir dan relevansi terhadap pengerjaan tugas akhir. Berdasarkan uraian pada bab ini, harapannya gambaran umum permasalahan dan pemecahan masalah pada tugas akhir dapat dipahami.

1.1 Latar Belakang Masalah

Penelitian ilmiah di era teknologi saat ini terus mengalami perkembangan yang sangat pesat. Hal ini dapat dilihat dari perkembangan penelitian yang terdadar di *website* repositori publikasi seperti *ScienceDirect*, *IEEE Xplore Digital Library*, dan lain-lain. Jumlah sitasi penelitian terdahulu pun menggambarkan pesatnya perkembangan penelitian di dunia.

Open Access Journal of Information Systems (OAJIS) merupakan repositori terbuka yang disediakan oleh Jurusan Sistem Informasi ITS untuk seluruh peneliti bidang sistem informasi di seluruh dunia. Dalam prakteknya, hasil pencarian data publikasi masih didasarkan pada kata kunci dari setiap dokumen. *Website OAJIS* pun hanya menyediakan kolom pencarian dengan langsung memasukkan kata kunci yang diinginkan. Belum ada kategorisasi lebih lanjut untuk memudahkan pencarian dan membantu pengguna untuk menemukan penelitian sejenis.

Masalah utama yang dihadapi *OAJIS* sendiri adalah sulitnya menentukan *reviewer* yang sesuai untuk mengevaluasi setiap penelitian yang masuk. Hal ini sangatlah penting, karena kesesuaian penelitian yang dievaluasi dengan bidang penelitian *reviewer* sangat berpengaruh pada hasil evaluasi.

Berdasarkan beberapa masalah di atas, penelitian ini menggunakan metode *Author-Topic Models* terhadap data

dokumen *OAJIS* sehingga dapat mengetahui topik-topik penelitian yang masuk ke repositori *OAJIS*. Harapan dari penelitian ini untuk mempermudah pengelola *OAJIS* dalam melakukan pengelompokkan dan rekomendasi penelitian kepada pengguna, juga mempermudah mereka untuk menemukan *reviewer* yang sesuai dengan penelitian baru yang masuk.

1.2 Perumusan Masalah

Rumusan masalah dari penelitian ini adalah:

1. Bagaimana melakukan *topic modeling* untuk menganalisa topik yang tersebar dalam data dokumen dari repositori *OAJIS* dengan metode *Author-Topic Models*?
2. Bagaimana perbandingan hasil *topic modeling* data tanpa *stemming*, dengan *stemming*, dan dengan *noun-part-of-speech tagged*?
3. Bagaimana mengevaluasi model untuk menilai performa prediksi model?

1.3 Batasan Masalah

Berdasarkan permasalahan yang disebutkan di atas, batasan masalah dalam penelitian ini adalah:

1. *Topic modeling* dilakukan pada data repositori publikasi *OAJIS* dengan menggunakan bahasa Indonesia.
2. Jenis data yang dianalisis menggunakan metode *Author-Topic Models* berupa data teks.
3. Data yang akan digunakan pada pengerjaan Tugas Akhir ini adalah data teks lengkap setiap publikasi yang tersimpan dalam repositori *OAJIS* sejak tahun 2013 hingga 2017.
4. Skenario yang diberikan pada data adalah tanpa *stemming*, dengan *stemming*, dan dengan *noun-part-of-speech tagged*.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Melakukan analisis *topic modeling* untuk mengetahui topik-topik penelitian di bidang sistem informasi apa saja yang tersimpan dalam publikasi *OAJIS*.
2. Melakukan analisis peminatan peneliti di bidang sistem informasi terkait publikasi yang tersimpan dalam repositori *OAJIS*.
3. Melakukan analisis model berdasarkan skenario tanpa *stemming*, dengan *stemming*, dan dengan *noun-part-of-speech tagged*.
4. Melakukan evaluasi model yang tepat dalam melakukan *topic modeling* menggunakan metode *Author-Topic Models*.
5. Melakukan pengujian untuk mengetahui kemampuan prediksi model dikaitkan dengan aplikasi model untuk memberikan rekomendasi *reviewer*.

1.5 Manfaat Penelitian

1. Bagi penulis, untuk mengetahui bagaimana *topic modeling* dapat menemukan topik-topik yang ada pada kumpulan data dokumen teks serta memahami implementasi metode *Author-Topic Models* dalam bidang penelitian Pengolahan Bahasa Alami (*Natural Language Processing*).
2. Bagi masyarakat, sebagai bentuk penelitian awal mengenai aplikasi metode *Author-Topic Models* dalam bahasa Indonesia agar dapat dikembangkan pada penelitian-penelitian selanjutnya.

1.6 Relevansi

Relevansi tugas akhir ini terhadap laboratorium Akuisisi Data dan Diseminasi Informasi (ADDI) adalah karena tugas akhir ini berkaitan dengan penerapan mata kuliah bidang keilmuan laboratorium ADDI. Mata kuliah tersebut antara lain Sistem Cerdas, Sistem Pendukung Keputusan, dan Penggalan Data dan Analitika Bisnis.

Halaman ini sengaja dikosongkan

BAB II TINJAUAN PUSTAKA

Bab ini akan membahas penelitian sebelumnya yang berhubungan dengan tugas akhir dan teori - teori yang berkaitan dengan permasalahan tugas akhir ini.

2.1 Studi Sebelumnya

Error! Reference source not found. menampilkan daftar penelitian sebelumnya yang mendasari tugas akhir ini.

Tabel 2.1.1 Studi sebelumnya

<i>1. The Author-Topic Model for Authors and Documents</i>
Penulis/Tahun/Sumber: Michal Rosen-Zvi, Thomas Griffiths, Mark Steyvers, Padhraic Smyth; 2004[1]
Metode: <i>Author-Topic Models</i>
Kesimpulan: <ul style="list-style-type: none">• Penelitian ini melakukan pengembangan metode <i>topic modelling Latent Dirichlet Allocation</i> menjadi sebuah <i>generative model</i> untuk dokumen yang dapat memuat informasi pengarang di dalamnya.• Pada <i>Author-Topic Models</i>, setiap pengarang diasosiasikan dengan distribusi multinomial melalui topik dan setiap topik juga diasosiasikan dengan distribusi multinomial melalui kata-kata.• <i>Author-Topic Models</i> menghasilkan peningkatan kemampuan prediksi yang signifikan. Hal ini dilihat dari nilai <i>perplexity</i> yang dibandingkan dengan <i>Author Model</i>, di mana minat penulis dimodelkan secara langsung sebagai distribusi probabilitas kata-kata.

<p>2. Probabilistic Author-Topic Models for Information Discovery</p>
<p>Penulis/Tahun/Sumber: Mark Steyvers, Padhraic Smyth, Michal Rosen-Zvi, Thomas Griffiths; 1999[2]</p>
<p>Metode: <i>Author-Topic Models for large text collections</i></p>
<p>Kesimpulan:</p> <ul style="list-style-type: none"> • Penelitian ini membuat aplikasi <i>Author-Topic Models</i> sehingga menghasilkan sebuah framework untuk eksplorasi, penemuan, dan <i>query-answering</i> dalam hubungan author dan topik pada korpora teks berskala besar. • Beberapa implementasi yang dilakukan pada data dari CiteSeer yang digunakan yaitu, mengetahui tren topik dari waktu ke waktu, pengelompokan <i>author</i> dan topik dari dokumen baru, dan mendeteksi paper yang paling berhubungan dan paling tidak berhubungan dengan satu <i>author</i>.
<p>3. Latent Dirichlet Allocation</p>
<p>Penulis/Tahun/Sumber: David M. Blei, Andrew Y. Ng, Michael I. Jordan; 2003[3]</p>
<p>Metode: <i>Latent Dirichlet Allocation (LDA)</i></p>
<p>Kesimpulan:</p> <ul style="list-style-type: none"> • Penelitian ini merepresentasikan sebuah dokumen direpresentasikan sebagai campuran acak atas topik yang belum terlihat, di mana setiap topik dikarakteristikan dengan distribusi kata-kata. • Implementasi <i>LDA</i> dilakukan untuk menyelesaikan beberapa masalah, yaitu <i>document modeling</i>, <i>document classification</i>, dan <i>collaborative filtering</i>.

- Pada document modeling, didapatkan nilai perplexity terendah yang dihasilkan oleh LDA, dibandingkan dengan pLSI, unigram model, dan mixture of unigrams.
- Pada *document classification*, dilakukan *training data* menggunakan *support vector machine (SVM)* dengan dua jenis data, yaitu data kata-kata yang sudah mengalami *dimensionality reduction* oleh LDA dan data seluruh kata yang ada. Hasil penelitian menunjukkan bahwa LDA model menunjukkan hasil klasifikasi kata yang lebih akurat pada data 50 topik.
- Pada *collaborative filtering*, performa LDA juga ditunjukkan oleh nilai *perplexity* yang lebih rendah dibandingkan dengan pLSI dan *mixture of unigrams*.

4. Discovering User Interest on Twitter with a Modified Author-Topic Model

Penulis/Tahun/Sumber:

Zhiheng Xu, Rong Lu, Liang Xiang, Qing Yang; 2011[4]

Metode:

- *Twitter-user Models*
- *Latent Dirichlet Allocation (LDA)*
- *Author-Topic Models*

Kesimpulan:

- Pada penelitian ini dilakukan sistematik modeling dari ketertarikan topik oleh pengguna twitter menggunakan metode *Author-Topic Models* yang telah disesuaikan dengan hubungan antara tweet dan penggunanya.
- Performa dari *Twitter-User Models* dibandingkan dengan LDA dan *Author-Topic Models* berdasarkan dua metric pengukuran, yaitu *perplexity of held-out content* dan *top words discovered* untuk setiap topik dan pengguna Twitter.
- Pada *Twitter-User Models*, kata-kata yang kurang berhubungan dengan minat pengguna sudah disaring terlebih dahulu, sehingga dapat memunculkan kata-kata

<p>yang lebih berhubungan dengan setiap topik, daripada <i>Author-Topics Models</i> dan <i>LDA</i>.</p>
<p>5. Mining Eclipse Developer Contributions via Author-Topic Models</p>
<p>Penulis/Tahun/Sumber: Erik Linstead, Paul Rigor, Sushil Bajracharya, Cristina Lopes, Pierre Baldi; 2007[5]</p>
<p>Metode: - <i>Author-Topic Models</i> - <i>Kullback-Leibler (KL) Divergence</i></p>
<p>Kesimpulan:</p> <ul style="list-style-type: none"> • Pada penelitian ini dilakukan implementasi <i>statistical author-topic models</i> pada <i>source code Eclipse 3.0</i> sehingga dapat menggali kontribusi dan kompetensi <i>developer</i> berdasarkan <i>source code</i> yang ada. • Penelitian ini juga menunjukkan <i>topic models</i> dapat menyediakan <i>developer similarity analysis</i> yang efektif dan berdasar statistik. • Penelitian ini menunjukkan bahwa algoritma <i>Author-Topic Models</i> menghasilkan automasi topik dan hasil kelompok topik yang masuk akal dan dapat dipahami.
<p>6. Building an Indonesian Rule-Based Part-of-Speech Tagger</p>
<p>Penulis/Tahun/Sumber: Fam Rashel, Andry Luthfi, Arawinda Dinakaramani, and Ruli Manurung; 2014 [6]</p>
<p>Metode: - <i>Rule-Based Approach</i></p>

Kesimpulan:

Penelitian ini membangun *part-of-speech tagger* yang dikhususkan untuk Bahasa Indonesia dengan menggunakan pendekatan *rule-based*. Sistem melakukan tokenisasi setiap kata pada dokumen dan mempertimbangkan kata majemuk serta *named-entity*-nya. Setiap token akan diberikan *tag* sesuai dengan kelas katanya. Saat ini akurasi sistem masih mencapai 79% berdasarkan *corpus* yang terdiri dari 250.000 token dan telah ditandai secara manual.

2.2 Dasar Teori**2.2.1 *Natural Language Processing***

Natural language processing (NLP) merupakan salah satu bidang riset dan implementasi studi yang mengeksplorasi bagaimana teknologi dapat digunakan untuk memahami dan memanipulasi bahasa alami, dalam bentuk teks maupun suara, untuk dapat dimanfaatkan dalam kehidupan sehari-hari[7]. Penelitian di bidang ini bertujuan untuk mengumpulkan pengetahuan tentang bagaimana perilaku manusia dalam memahami dan menggunakan bahasa. Berdasarkan hal tersebut, kemudian dikembangkan alat maupun teknik yang sesuai untuk membuat suatu sistem komputer yang dapat memahami bahasa alami manusia dan melakukan tugas yang diinginkan.

Dalam menjelaskan metode penyajian mengenai apa yang sebenarnya terjadi dalam *NLP*, terdapat suatu pendekatan yang disebut dengan '*level of language*' atau tingkat bahasa [8]. Yang dimaksud dengan tingkat bahasa adalah bahwa makna dari setiap tingkat bahasa berbeda, pun manusia menggunakan semua tingkat bahasa untuk memperoleh pemahaman yang lebih baik. Semakin banyak suatu sistem *NLP* dapat memfasilitasi tingkat bahasa yang ada, maka semakin baik. Beberapa tingkat bahasa tersebut adalah *phonology* (interpretasi kata dalam suara), *morphology* (susunan kata dan bentuknya), *lexical* (pemahaman arti setiap kata secara individu), *syntactic* (analisa kata-kata untuk mengetahui struktur tata bahasa),

semantic (interaksi arti antar kata dalam kalimat dalam konteks), *discourse* (pemahaman teks secara keseluruhan dengan menghubungkan komponen kalimat), dan *pragmatic* (penggunaan bahasa dan konteks isi teks untuk dipahami) [8]

Dalam implementasinya, *NLP* dapat diaplikasikan dalam berbagai permasalahan [8], di antaranya;

- *Information Retrieval* adalah salah satu aplikasi dari *NLP* yang berfokus pada representasi, penyimpanan, pengelolaan, dan akses pada item-item informasi. Item informasi disini dapat berupa dokumen, paragraf, halaman web, gambar, dokumen lisan, video, dan lain-lain.[9]
- *Information Extraction* yaitu salah satu bidang *NLP* yang berfokus pada *recognition*, *tagging*, dan ekstraksi informasi menjadi representasi elemen penting dari informasi secara terstruktur dari data teks berskala besar.
- *Question-Answering* berkebalikan dengan *information retrieval*, dimana IR memberikan daftar dokumen yang berhubungan dalam merespon pertanyaan pengguna, *question-answering* akan memberikan jawaban dari pertanyaan pengguna secara langsung.
- *Summarization* merupakan kelanjutan dari tingkat bahasa dalam *level discourse*, sehingga dapat mengurangi teks berskala besar menjadi lebih ringkas dengan representasi dokumen asli.
- *Machine Translation* merupakan salah satu dari implementasi *NLP* yang tertua, yaitu dengan mengartikan suatu teks bahasa yang satu menjadi bahasa yang lain.
- *Named-Entity Recognition*[10] menjelaskan tentang aliran teks, menentukan item yang terdapat di dalam teks yang berhubungan dengan nama tertentu, seperti nama tempat maupun nama orang, dan tipe dari setiap nama atau tempat yang dimaksud.

2.2.2 *Topic Modeling*

Topic modeling merupakan algoritma dengan metode statistik yang menganalisa kata-kata dari teks asli untuk mengetahui tema yang melingkupi dokumen terkait, bagaimana tema-tema yang ada saling berhubungan, juga bagaimana tema-tema tersebut berubah seiring dengan berjalannya waktu [11]. Algoritma ini tidak membutuhkan tahap anotasi atau penerjemahan dan pemberian label pada dokumen terlebih dahulu, karena topik-topik dari dokumen didapatkan berdasarkan analisis yang dilakukan pada teks asli. Dengan melakukan *topic modeling*, pengelolaan dan peringkasan arsip elektronik dalam skala besar yang sulit dilakukan maupun diterjemahkan oleh manusia secara manual akan menjadi lebih mudah.

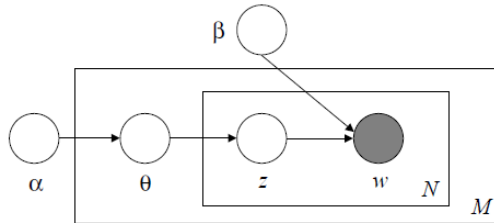
Dimensi dari teks pun akan mengecil sehingga dapat mempermudah proses penyimpanan dan penggalian informasi dari teks maupun dokumen terkait. Tidak hanya teks, metode ini juga dapat diadaptasi untuk berbagai jenis data sehingga telah banyak digunakan untuk mengetahui pola-pola persebaran data dalam berbagai masalah, seperti data genetik, gambar, jaringan sosial, dan lain-lain.

2.2.3 *Latent Dirichlet Allocation (LDA)*

Latent Dirichlet Allocation (LDA) merupakan sebuah topik model paling sederhana yang menggunakan model probalistik untuk suatu kumpulan data diskrit, seperti korpora teks. *LDA* adalah sebuah *three-level hierarchical Bayesian model*, dimana setiap item dalam kumpulan data dimodelkan sebagai suatu campuran kata-kata yang terbatas dari suatu kumpulan topik [3]. Metode ini bertujuan untuk mendapatkan ringkasan mengenai anggota dari koleksi data sehingga memungkinkan proses yang efisien dari data berskala besar. Hal ini berguna untuk tugas mendasar seperti klasifikasi, pendeteksian kesalahan, *summarization*, dan penilaian kesamaan dan relevansi data. Secara singkat, dengan menggunakan metode

LDA dapat diketahui topik-topik yang terdapat dalam suatu dokumen berdasarkan persebaran kata-kata yang ada.

Pada penelitiannya, Blei merepresentasikan model dari *LDA* seperti pada Gambar 2.2.1.



Gambar 2.2.1 Model LDA

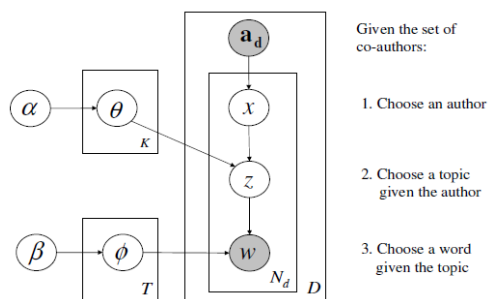
Pada ilustrasi di atas dapat dilihat bahwa terdapat tiga tingkatan *Bayesian model* yang menyusun metode *LDA*. Kotak *M* merepresentasikan dokumen yang dianalisa, sedangkan kotak *N* merepresentasikan perulangan pilihan topik-topik dan kata-kata dalam sebuah dokumen. Parameter α dan β merupakan parameter distribusi topik pada tingkatan *corpus*. Parameter α akan menentukan distribusi topik-topik yang ada pada dokumen, sedangkan parameter β menentukan distribusi kata-kata dalam topik. Semakin besar nilai α , maka jumlah topik yang dibahas dalam dokumen semakin banyak. Semakin besar nilai β , menggambarkan distribusi kata-kata dalam topik yang semakin banyak dan cakupan topik yang semakin luas. Sebaliknya, semakin kecil nilai β , menandakan distribusi kata-kata pada topik yang semakin sedikit, sehingga menandakan bahwa topik lebih spesifik. Parameter θ merupakan parameter tingkat dokumen yang menggambarkan persebaran topik pada dokumen tertentu. Lalu parameter z dan w merupakan parameter tingkat kata. Parameter z menggambarkan topik yang terhubung dengan kata tertentu, sedangkan parameter w menggambarkan kata yang berhubungan dengan topik tertentu.

2.2.4 Author-Topic Models

Author-Topic Models merupakan sebuah metode perpanjangan dari *LDA*, dengan tambahan integrasi informasi mengenai

author suatu dokumen pada topic modeling [4]. Pada author-topic models diasumsikan bahwa setiap author pada setiap kumpulan dokumen direpresentasikan dengan distribusi topik dan setiap kata diasosiasikan dengan dua variabel tersembunyi, yaitu author dan topik [2]. Berbeda dengan proses pada LDA, author-topic model akan memilih seorang author dari daftar author, kemudian menentukan topik dari distribusi topik yang telah terasosiasi dengan author yang telah dipilih tadi. Dari topik yang ada, akan dipilih kata-kata yang berkaitan dengan topik berdasarkan distribusi kata pada topik terkait.

Pada penelitiannya, Steyvers (2004) mendefinisikan *Author-Topic Models* sesuai Gambar 2.2.2.



Gambar 2.2.2 Model Author-Topic Models

Pada **Error! Reference source not found.** diketahui bahwa susunan model mirip dengan *LDA*, dengan tambahan parameter *author*. Kotak K merupakan kumpulan *author* yang diasosiasikan dengan distribusi multinomial melalui topik yang ada, direpresentasikan dengan parameter θ . Oleh karena itu, pada gambar 2.2 terlihat bahwa θ berhubungan dengan variabel z , di mana z menggambarkan topik pada dokumen. Selanjutnya, kotak T merupakan kumpulan topik yang juga diasosiasikan dengan distribusi multinomial melalui kata-kata, direpresentasikan dengan parameter ϕ . Pada gambar dapat dilihat bahwa ϕ berhubungan dengan variabel w , dimana w merupakan kata pada dokumen. Pada setiap kata dalam dokumen, dipilih sebuah *author* x dari A_d , lalu topik z ditentukan berdasarkan distribusi θ yang terhubung dengan

author x . Kemudian kata w juga ditentukan berdasarkan distribusi ϕ yang terhubung dengan topik z . Proses ini terus berulang selama N kali untuk membentuk sebuah dokumen d .

Author-Topic Models ini merupakan salah satu metode yang dapat menyederhanakan proses penulisan dokumen ilmiah menjadi serangkaian langkah probabilistik sederhana. *Author-Topic Models* tidak hanya menemukan topik-topik yang menggambarkan suatu dokumen, tetapi juga dapat membantu menentukan *author* mana saja yang berhubungan dengan setiap topik yang ada.

Beberapa aplikasi dari *Author-Topic Models* yang dapat diimplementasikan untuk menyelesaikan permasalahan [2] yaitu untuk mengetahui tren topik terkini dari waktu ke waktu, pendeteksian dokumen secara otomatis berdasarkan *author*, dan penentuan topik dan *author* untuk dokumen baru yang tidak terdapat dalam kumpulan dokumen yang dimiliki.

2.2.5 *Part-of-Speech Tag*

Part-of-Speech tag merupakan kategori tata bahasa untuk kata-kata seperti kata benda, kata kerja, kata sifat, kata keterangan, dan lain-lain [6]. *Part-of-Speech tagging* merupakan proses pemberian *tag* berdasarkan *part-of-speech tag* pada kata-kata dalam kalimat atau teks dokumen. Sistem ini merupakan alat yang biasa digunakan dalam banyak aplikasi pemrosesan bahasa, seperti *word sense disambiguation*, *parsing*, *question answering*, dan *machine translation* [12]. Contoh penggunaan *part-of-speech tag* pada sebuah kalimat dapat dilihat pada Gambar 2.2.3 berikut.

<u>Budi</u>	<u>makan</u>	<u>nasi</u>	<u>di</u>	<u>rumah</u>	<u>.</u>
NNP	VB	NN	IN	NN	Z

Gambar 2.2.3 Contoh hasil *tagging* menggunakan POS Tag Indonesia

2.2.6 Evaluasi Model Berdasarkan Nilai *Perplexity*

Untuk melakukan penilaian pada model yang telah dihasilkan, dapat digunakan suatu pengukuran yang sudah sering digunakan dalam bidang *topic modeling*, yaitu nilai *perplexity*. *Perplexity* merupakan suatu standar pengukuran yang dapat digunakan untuk menilai performa dari suatu model probabilistik [1]. Performa model yang baik ditunjukkan dengan nilai *perplexity* yang kecil [2]. Semakin kecil nilai *perplexity*, semakin jelas berbeda topik-topik yang ditemukan oleh model. Sebaliknya, semakin besar nilai *perplexity*, menunjukkan semakin miripnya topik-topik yang ditemukan berdasarkan distribusi kata-kata yang ada pada dokumen.

Nilai *perplexity* dari sebuah kelompok kata-kata uji didefinisikan sesuai persamaan (1).

$$perplexity(w_d|a_d) = \exp \left[-\frac{\ln p(w_d|a_d)}{N_d} \right] \quad (1)$$

2.2.7 Evaluasi Model Berdasarkan *Topic Coherence*

Perplexity merupakan evaluasi model yang telah sering digunakan dalam *topic modeling*, namun pada penelitian sebelumnya diketahui bahwa metode ini sering kali berkorelasi secara negatif dengan kualitas topik berdasarkan pendapat manusia [13], sedangkan *topic coherence* didasarkan pada penelitian yang memfokuskan pada kualitas topik berdasarkan pendapat manusia yang mengikutsertakan interpretasi manusia [14]. Berdasarkan studi ini, *topic coherence* merupakan salah satu bentuk evaluasi model yang dapat memberikan penilaian terhadap kualitas topik yang sesuai dengan sudut pandang manusia. Semakin tinggi nilai *topic coherence*, berarti model menghasilkan kelompok-kelompok kata dalam topik yang semakin baik. Berdasarkan nilai *topic coherence*, dapat pula dilakukan perhitungan untuk dapat mengurutkan topik terbaik yang dibentuk oleh model, sehingga diketahui topik mana yang benar-benar berbeda dengan topik lainnya dan memiliki kualitas yang sesuai dengan pendapat manusia.

Nilai *topic coherence* dari kelompok kata-kata dalam topik didefinisikan sesuai persamaan (2).

$$Score_{UMass}(w_i, w_j) = \log \frac{D(w_i, w_j) + 1}{D(w_i)} \quad (2)$$

2.2.8 Perhitungan Nilai *Similarity Matrix* berdasarkan *Hellinger Distance*

Hellinger Distance merupakan salah satu pengukuran jarak dalam ilmu statistik dan probabilitas yang digunakan untuk mengkuantisasi similaritas di antara dua distribusi probabilitistik. Dalam aplikasinya, *Hellinger distance* lebih tepat untuk digunakan dengan asosiasi antara kata-kata dan dokumen. Hal ini disebabkan oleh nilainya yang tidak negatif dan memiliki hasil yang lebih baik untuk pendekatan probabilitistik [15]. Berdasarkan *Shunzhi Zhu*, hasil dari penggunaan *Hellinger Distance* lebih baik dalam melakukan *information retrieval* jika dibandingkan dengan *Euclidean Distance*.

2.2.9 *Open Access Journal of Information Systems (OAJIS)*

Open Access Journal of Information Systems (OAJIS) merupakan sebuah repositori publikasi yang dapat diakses oleh siapapun, berisikan berbagai publikasi jurnal nasional dan internasional. *OAJIS* sendiri dikelola oleh AISINDO dan Jurusan Sistem Informasi ITS. Tujuan dari *OAJIS* adalah untuk menyebarkan pengetahuan yang dikhususkan mengenai topik sistem informasi, bekerja sama dengan ilmuwan, akademisi, professional, dan praktisi untuk menyediakan akses gratis ke seluruh dunia. Fokus penelitian di *OAJIS* sendiri adalah isu dan perkembangan di dunia sistem informasi.

BAB III METODOLOGI

Bab ini menjelaskan tentang metodologi yang akan digunakan dalam penyusunan tugas akhir. Metodologi akan digunakan sebagai panduan dalam penyusunan tugas akhir agar terarah dan sistematis.

3.1 Tahapan Pelaksanaan Tugas Akhir

Alur tahapan pelaksanaan yang dilakukan dalam mengerjakan tugas akhir ini sesuai dengan Diagram 3.1.1.

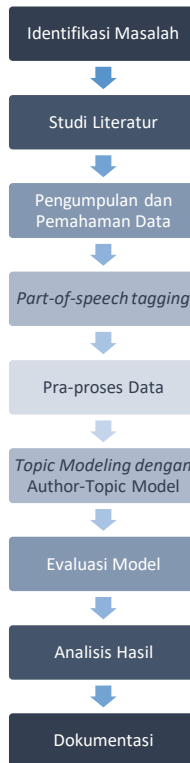


Diagram 3.1.1 Tahapan pelaksanaan tugas akhir

3.1.1 Identifikasi Masalah

Pada tahap ini dilakukan identifikasi masalah terkait studi kasus yaitu *Open Access Journal of Information Systems (OAJIS)*. Identifikasi masalah dilakukan dengan memahami proses bisnis dan pengelolaan data pada *OAJIS* sehingga didapatkan beberapa masalah yang terjadi. Dari tahap identifikasi masalah ini akan didapatkan permasalahan yang terjadi pada sistem *OAJIS* dan penentuan metode yang tepat untuk dapat menyelesaikan masalah. Beberapa masalah yang ditemui adalah sulitnya menentukan *reviewer* penelitian yang sesuai dengan kepakarannya, karena belum adanya penyajian data publikasi yang didasarkan pada pengelompokan topik publikasi. Hal ini juga sangat dibutuhkan ketika pengguna sedang melakukan pencarian suatu publikasi tertentu dan mengalami kesulitan saat menentukan kata kunci atau menemukan penelitian sejenis. Berdasarkan kondisi tersebut, dirasa perlu untuk melakukan identifikasi topik dari setiap dokumen publikasi sehingga akan memudahkan proses pengelompokan dan prediksi topik.

3.1.2 Studi Literatur

Pada tahap ini dilakukan studi literatur sesuai dengan permasalahan yang ditemui serta metode yang dibutuhkan untuk menyelesaikan permasalahan yang ada. Studi literatur yang akan dilakukan adalah pembelajaran mengenai penelitian terdahulu terkait masalah yang telah ditemui, yaitu tentang penggalian informasi dan penentuan topik dari data dokumen teks, penelitian terdahulu mengenai *topic modeling* dan penggalian data, dan algoritma serta metode yang dapat digunakan untuk menyelesaikan masalah yang ada secara tepat. Literatur utama yang digunakan dalam tugas akhir ini adalah penelitian dengan judul *Learning Author-Topic Models from Text Corpora* [16], penelitian berjudul *The Author-Topic Models* [17] dalam aplikasinya menggunakan *Gensim*, dan implementasi *Author-Topic Models* pada *NIPS dataset* dalam *The Author-Topic Model: LDA with Metadata* [18] yang dikeluarkan oleh *Gensim*.

3.1.3 Proses Pengumpulan Data

Pada tahap ini dilakukan proses pengumpulan dan pemahaman data yang diperlukan untuk menyelesaikan permasalahan yang ada. Data didapatkan dari repositori *OAJIS* Jurusan Sistem Informasi ITS dalam bentuk *database* dengan format *.xlsx* dan *file* setiap penelitian dalam bentuk *pdf* untuk sumber data teks lengkapnya. Data yang dikumpulkan merupakan data periodik dari tahun 2013 hingga 2017. Atribut dari data pada *database* adalah nama *file*, judul, abstrak Indonesia, *author*, dan *keyword* Indonesia. Contoh data yang didapatkan dari repositori *OAJIS* ditunjukkan pada Tabel 3.1.1.

Tabel 3.1.1 Contoh data dari repositori *OAJIS*

nama file	Judul	Abstrak Indonesia	Author	keywords indonesia
R01	E-SUPPLY CHAIN MANAGEMENT UNTUK MEMAJUKAN USAHA KECIL MENENGAH DALAM STRATEGI PERSAINGAN BISNIS	Perusahaan membutuhkan sebuah sistem informasi yang dapat mengelola hubungan antara perusahaan dengan pelanggan dan mengaplikasikan konsep customer relationship management agar dapat dimanfaatkan pada customer care center department di perusahaan.	Afdhal Syafnur, Rika Nofitri, Winda Sulastri	Sistem, Informasi, Supply, Management (SCM).

Sedangkan contoh data dalam bentuk *.pdf* terdapat pada Gambar 3.1.1. Data dalam bentuk *.pdf* ini kemudian dikonversi menjadi bentuk *.txt* untuk memudahkan proses data.

TATA KELOLA INTEGRASI SISTEM INFORMASI PT.X DENGAN MENGGUNAKAN FRAMEWORK COBIT 4.1

Budiya Surya Putra

Sistem Informasi, Fakultas Teknik, Institut Informatika Indonesia

Jl.Raya Sukomanunggal Jaya No.3, Surabaya, 60188

Telp : (031) 7346375, Fax (031) 7349324

E-mail : budiya@iiti.ac.id

Abstrak

Sistem informasi yang terintegrasi membutuhkan tata kelola dikarenakan memberikan pengarahan dan memastikan pencapaian kinerja sesuai dengan tujuan yang diinginkan. Sehingga TI menjadi searah dengan perusahaan dan manfaat yang dijanjikan dapat terealisasi. Selain itu Sistem yang ada juga dapat dipertanggung jawabkan, sehingga dapat mengurangi resiko yang akan terjadi di kemudian hari. Dengan menggunakan tata kelola, kesalahan – kesalahan yang terjadi pada Sistem Informasi dapat diminimalisasikan. Sehingga Sistem Informasi yang dibuat dapat dijalankan sesuai dengan tujuan. Sehingga perawatan dan manajemen resiko juga dapat dilakukan.

Kata kunci : tata Kelola, COBIT, integrasi sistem informasi.

Gambar 3.1.1 Contoh file paper dalam bentuk .pdf

Setelah dilakukan konversi *file* dari bentuk *.pdf* menjadi *.txt*, dilakukan pembersihan data setiap *file* dengan hanya mengambil isi *paper* mulai dari abstrak Bahasa Indonesia hingga kesimpulan, sebelum daftar pustaka. Hal ini dilakukan karena teks yang dianggap relevan untuk digunakan sebagai data permodelan hanya isi *paper* mulai dari abstrak Bahasa Indonesia hingga kesimpulan. Teks di luar itu hanya akan membiarkan data yang digunakan dan kurang relevan karena terdapat *header* dan *footer* jurnal, judul paper, alamat penulis, *e-mail*, dan lain-lain.

Contoh salah satu *file .txt* yang telah dibersihkan dapat dilihat pada Gambar 3.1.2.

Sistem informasi yang terintegrasi membutuhkan tata kelola dikarenakan memberikan pengarahan dan memastikan pencapaian kinerja sesuai dengan tujuan yang diinginkan. Sehingga TI menjadi searah dengan perusahaan dan manfaat yang dijanjikan dapat terealisasi. Selain itu Sistem yang ada juga dapat dipertanggung jawabkan, sehingga dapat mengurangi resiko yang akan terjadi di kemudian hari. Dengan menggunakan tata kelola, kesalahan kesalahan yang terjadi pada Sistem Informasi dapat diminimalisasikan. Sehingga Sistem Informasi yang dibuat dapat dijalankan sesuai dengan tujuan. Sehingga perawatan dan manajemen resiko juga dapat dilakukan. Kata kunci : tata Kelola, COBIT, integrasi sistem informasi. Abstract Information integration system requires governance due to provide guidance and ensure the achievement of performance in accordance with the desired goals. So as to be in line with the company's IT and promised benefits can be realized. In addition there is also a system that can be justified, so as to reduce the risk that will happen in the future. By using governance, errors-errors that occur on Information Systems can be minimized. Information System so created can be run according to the purpose. So that care and risk management can also be done. Keyword : IT governance, COBIT, information integration system. 1. PENDAHULUAN Sistem informasi yang terintegrasi membutuhkan tata kelola dikarenakan memberikan pengarahan dan memastikan pencapaian kinerja sesuai dengan tujuan yang diinginkan [1]. Selain itu, ada beberapa alasan mengapa tata kelola diperlukan dalam pembuatan sistem informasi, yaitu : Tatakelola diperlukan untuk memberikan pengarahan dan memastikan pencapaian kinerja sesuai dengan tujuan yang diinginkan. TI menjadi searah dengan perusahaan dan manfaat

Gambar 3.1.2 Contoh data file paper yang telah dikonversi menjadi .txt

3.1.4 Part-of-Speech Tag

Sebelum masuk ke tahap pra-proses data, dilakukan *part-of-speech tagging* pada data sebagai persiapan data dengan skenario *noun part-of-speech tag*. Skenario ini dilakukan berdasarkan hipotesa kata-kata yang menyusun topik penelitian lebih relevan pada kata benda dibandingkan dengan kata kerja. Oleh karena itu, dengan hanya mengambil kata benda, akan menghapus kata kerja dari teks data. Tahap ini dilakukan dengan menggunakan POS Tag Indonesia [6] (<http://bahasa.cs.ui.ac.id/postag/>) dari Laboratorium Perolehan Informasi Fakultas Ilmu Komputer, Universitas Indonesia. Sistem ini berbasis bahasa *Perl* dan harus dijalankan dalam sistem operasi berbasis UNIX.

Contoh hasil *part-of-speech tagging* yang dilakukan dengan menggunakan POS Tak Indonesia dapat dilihat pada Gambar 3.1.3. Dari hasil tersebut, token yang bukan kata benda (NN) akan dihapus karena yang akan digunakan sebagai data hanyalah kata benda.

Sistem informasi	NN
yang	SC
terintegrasi	NN
membutuhkan	VB
tata kelola	NN
dikarenakan	X
memberikan	VB
pengarahan	NN
dan	CC
memastikan	VB
pencapaian	NN
kinerja	NN
sesuai	JJ
dengan	SC,IN
tujuan	NN
yang	SC
diinginkan	VB
.	Z

Gambar 3.1.3 Hasil *tagging* menggunakan POS Tag Indonesia

3.1.5 Pra-proses Data

Sebelum melakukan *topic modeling* pada suatu data, dibutuhkan tahap untuk mengolah dan membersihkan data sehingga data siap digunakan sebagai masukkan untuk proses *topic modeling*. Pada tahapan ini, terdapat lima langkah yang dilakukan untuk membersihkan data yang akan digunakan,

yaitu *data cleaning*, *case folding*, *stemming*, *stop-word removal*, dan *tokenization*.

3.1.5.1 Data Cleaning

Pada tahap *data cleaning*, dilakukan pembersihan data dari karakter numerik dan simbol-simbol. Selain itu, dilakukan juga penghapusan satu digit karakter. Hal ini dilakukan untuk menghindari tingginya frekuensi kemunculan karakter numerik, simbol, dan satu digit karakter yang luput dari proses pembersihan data untuk mengambil isi abstrak hingga kesimpulan yang telah dilakukan sebelum pra-proses data. Contoh tahapan *data cleaning* dapat dilihat pada Tabel 3.1.2.

Tabel 3.1.2 Data cleaning

Sebelum <i>data cleaning</i>	Setelah <i>data cleaning</i>
Keyword : IT governance COBIT information integration system 1. PENDAHULUAN Sistem informasi yang terintegrasi membutuhkan tata kelola dikarenakan memberikan pengarahan dan memastikan pencapaian kinerja sesuai dengan tujuan yang diinginkan [1]. Selain itu, ada beberapa alasan mengapa tata kelola diperlukan dalam pembuatan sistem informasi, yaitu : Tatakelola diperlukan untuk memberikan pengarahan dan memastikan pencapaian kinerja sesuai dengan tujuan yang diinginkan.	Keyword IT governance COBIT information integration system PENDAHULUAN Sistem informasi yang terintegrasi membutuhkan tata kelola dikarenakan memberikan pengarahan dan memastikan pencapaian kinerja sesuai dengan tujuan yang diinginkan Selain itu ada beberapa alasan mengapa tata kelola diperlukan dalam pembuatan sistem informasi yaitu Tatakelola diperlukan untuk memberikan pengarahan dan memastikan pencapaian kinerja sesuai dengan tujuan yang diinginkan

3.1.5.2 Case Folding

Pada langkah ini, dilakukan perubahan keseluruhan teks menjadi huruf kecil (*lowercase*). Hal ini dilakukan untuk menghindari perbedaan interpretasi mesin pada saat

menemukan kata yang sama dengan jenis huruf yang berbeda. Contoh perubahan yang terjadi pada *case folding* dapat dilihat pada Tabel 3.1.3.

Tabel 3.1.3 Case folding

Sebelum <i>case folding</i>	Setelah <i>case folding</i>
Keyword IT governance COBIT information integration system PENDAHULUAN Sistem informasi yang terintegrasi membutuhkan tata kelola dikarenakan memberikan pengarahan dan memastikan pencapaian kinerja sesuai dengan tujuan yang diinginkan Selain itu ada beberapa alasan mengapa tata kelola diperlukan dalam pembuatan sistem informasi yaitu Tatakelola diperlukan untuk memberikan pengarahan dan memastikan pencapaian kinerja sesuai dengan tujuan yang diinginkan	keyword it governance cobit information integration system pendahuluan sistem informasi yang terintegrasi membutuhkan tata kelola dikarenakan memberikan pengarahan dan memastikan pencapaian kinerja sesuai dengan tujuan yang diinginkan selain itu ada beberapa alasan mengapa tata kelola diperlukan dalam pembuatan sistem informasi yaitu tatakelola diperlukan untuk memberikan pengarahan dan memastikan pencapaian kinerja sesuai dengan tujuan yang diinginkan

3.1.5.3 Stemming

Pada langkah ini dilakukan penggantian seluruh kata-kata yang bukan kata dasar menjadi kata dasar tanpa diikuti imbuhan. Hal ini dilakukan untuk menyeragamkan kata-kata yang bermakna sama, sehingga mesin tidak salah menginterpretasikan kata-kata bermakna sama yang berbeda karena keberadaan imbuhan. Dalam melakukan *stemming*, digunakan *library Stemmer* bahasa Indonesia. *Library* tersebut adalah *library Sastrawi* yang berlisensi *MIT* yang menggunakan kamus kata dasar dari Kateglo dalam aplikasinya. Contoh langkah *stemming* dapat dilihat pada Tabel 3.1.4.

Tabel 3.1.4 Stemming

Sebelum <i>stemming</i>	Setelah <i>stemming</i>
keyword it governance cobit information integration system	keyword it governance cobit information integration system

<p>pendahuluan sistem informasi yang terintegrasi membutuhkan tata kelola dikarenakan memberikan pengarahan dan memastikan pencapaian kinerja sesuai dengan tujuan yang diinginkan selain itu ada beberapa alasan mengapa tata kelola diperlukan dalam pembuatan sistem informasi yaitu tatakelola diperlukan untuk memberikan pengarahan dan memastikan pencapaian kinerja sesuai dengan tujuan yang diinginkan</p>	<p>dahulu sistem informasi yang integrasi butuh tata kelola karena beri arah dan pasti capai kinerja sesuai dengan tujuan yang ingin selain itu ada beberapa alasan mengapa tata kelola perlu dalam buat sistem informasi yaitu tatakelola perlu untuk beri arah dan pasti capai kinerja sesuai dengan tujuan yang ingin</p>
--	--

3.1.5.4 *Stopwords removal*

Pada langkah ini dilakukan penghapusan *stopwords* yang terdapat dalam teks. *Stopwords* adalah kata-kata yang sering muncul secara umum namun dianggap tidak memiliki makna seperti kata kerja, kata sifat dan kata benda. Contoh kata-kata *stopwords* adalah di, pada, dari, dengan, serta, atau, juga, namun, memang, dan lain-lain. Contoh langkah penghapusan *stopwords* dapat dilihat pada Tabel 3.1.5.

Tabel 3.1.5 *Stopwords removal*

Sebelum penghapusan <i>stopwords</i>	Setelah penghapusan <i>stopwords</i>
<p>keyword it governance cobit information integration system dahulu sistem informasi yang integrasi butuh tata kelola karena beri arah dan pasti capai kinerja sesuai dengan tujuan yang ingin selain itu ada beberapa alasan mengapa tata kelola perlu dalam buat sistem informasi yaitu tatakelola perlu untuk beri arah dan pasti capai</p>	<p>keyword it governance cobit information integration system sistem informasi integrasi butuh tata kelola beri arah pasti capai kinerja tujuan ingin alasan mengapa tata kelola perlu buat sistem informasi tatakelola perlu beri arah pasti capai kinerja sesuai tujuan ingin</p>

kinerja sesuai dengan tujuan yang ingin	
---	--

3.1.5.5 Tokenisasi

Pada langkah ini dilakukan tokenisasi, yaitu pemotongan atau pemisahan kalimat teks menjadi potongan kata tunggal yang didasarkan pada spasi sebagai batas pemisahannya. Contoh langkah tokenisasi dapat dilihat pada Tabel 3.1.6.

Tabel 3.1.6 Tokenisasi

keyword	it	governance	cobit	information
integration	system	sistem	informasi	integrasi
butuh	tata	kelola	beri	arah
pasti	capai	kinerja	tujuan	ingin
alasan	mengapa	tata	kelola	perlu
buat	sistem	informasi	tatakelola	perlu

3.1.6 Topic Modeling dengan Author-Topic Model

Pada tahap ini dilakukan pembentukan model dengan metode *Author-Topic Models*. Metode ini dimulai dengan memetakan kata-kata secara acak pada beberapa topik dan *author* dari kumpulan *author* dan topik yang terdapat pada dokumen. Setelah itu dilakukan iterasi hingga jumlah tertentu untuk mencapai konvergen sehingga didapatkan beberapa hasil sampel pengelompokan topik dan *author* berdasarkan algoritma yang dijalankan.

TOPIC 4		TOPIC 13		TOPIC 28		TOPIC 9	
WORD	PROB.	WORD	PROB.	WORD	PROB.	WORD	PROB.
LIGHT	.0305	RECOGNITION	.0500	KERNEL	.0547	SOURCE	.0389
RESPONSE	.0282	CHARACTER	.0334	VECTOR	.0293	INDEPENDENT	.0376
INTENSITY	.0252	TANGENT	.0246	SUPPORT	.0293	SOURCES	.0344
RETINA	.0241	CHARACTERS	.0232	MARGIN	.0239	SEPARATION	.0322
OPTICAL	.0233	DISTANCE	.0197	SVM	.0196	INFORMATION	.0319
KOCH	.0190	HANDWRITTEN	.0166	DATA	.0165	ICA	.0276
BACKGROUND	.0162	DIGITS	.0154	SPACE	.0161	BLIND	.0227
CONTRAST	.0145	SEGMENTATION	.0142	KERNELS	.0160	COMPONENT	.0226
CENTER	.0124	DIGIT	.0124	SET	.0146	SEJNOWSKI	.0224
FEEDBACK	.0118	IMAGE	.0111	MACHINES	.0132	NATURAL	.0183
AUTHOR	PROB.	AUTHOR	PROB.	AUTHOR	PROB.	AUTHOR	PROB.
Koch_C	.0903	Simard_P	.0602	Scholkopf_B	.0774	Sejnowski_T	.0627
Boahen_K	.0320	Martin_G	.0340	Smola_A	.0685	Bell_A	.0378
Skrzyppek_J	.0283	LeCun_Y	.0339	Vapnik_V	.0487	Yang_H	.0349
Liu_S	.0250	Henderson_D	.0289	Burges_C	.0411	Lee_T	.0348
Delbruck_T	.0232	Denker_J	.0245	Ratsch_G	.0296	Atlas_H	.0290
Etienne-C_R	.0210	Revow_M	.0206	Mason_L	.0232	Parra_L	.0271
Bair_W	.0178	Rashid_M	.0205	Platt_J	.0225	Cichocki_A	.0262
Bialek_W	.0133	Rumelhart_D	.0185	Cristianini_N	.0178	Hyvarinen_A	.0242
Yasui_S	.0106	Sackinger_E	.0181	Laskov_P	.0160	Amari_S	.0160
Hsu_K	.0103	Flann_N	.0142	Chapelle_O	.0152	Oja_E	.0143

Gambar 3.1.4 Sampel hasil *topic modeling*

Pada Gambar 3.1.4 dapat dilihat sampel hasil dari *topic modeling* yang dilakukan oleh metode *Author-Topic Model*. Beberapa sampel ini diambil untuk membandingkan perbedaan setiap hasil yang ada, sehingga dapat dilihat perilaku pengelompokan topik yang terjadi.

3.1.7 Evaluasi Model

Setelah melakukan eksperimen terhadap bentuk model yang dihasilkan, dilakukan evaluasi untuk mengetahui kemampuan model dalam melakukan prediksi penentuan topik penelitian *author* berdasarkan distribusi kata-kata dan *author* publikasi. Evaluasi model secara kuantitatif dilakukan dengan mengukur nilai *perplexity* dan *topic coherence* setiap model. Kemudian untuk melakukan pengujian, dilakukan perhitungan kemiripan topik penelitian dokumen dengan *author* yang ada berdasarkan model yang telah terbentuk. Perhitungan ini dilakukan berdasarkan nilai *similarity* menggunakan *Hellinger Distance*.

3.1.8 Analisis Hasil

Pada tahap ini dilakukan analisa hasil setelah mengetahui model terbaik yang didapatkan dari hasil eksperimen dan pengujian. Pada hasil validasi dan pengujian model berdasarkan topik yang didapatkan, dianalisa hubungan antar topik beserta kesesuaian antara *author* dengan topik, serta kemampuan topik dalam mengenali kesesuaian *author* dengan dokumen yang ada berdasarkan nilai probabilitas distribusi kata-kata pada dokumen dengan distribusi kata-kata pada setiap topik.

3.1.9 Dokumentasi

Pada tahap ini dilakukan pembuatan dokumentasi proses pengerjaan tugas akhir untuk menghasilkan buku tugas akhir. Harapannya, buku tugas akhir ini dapat dimanfaatkan sebagai referensi untuk penelitian selanjutn

BAB IV PERANCANGAN

Bab ini menjelaskan bagaimana rancangan dari penelitian tugas akhir yang meliputi subyek dan obyek dari penelitian, pemilihan subyek dan obyek penelitian dan bagaimana penelitian ini akan dilakukan.

4.1 Pengambilan Data

Untuk dapat menganalisa topik informasi yang diambil dari isi penelitian repositori Jurusan Sistem Informasi, yaitu *Open Access Journal of Information Systems*, dibutuhkan data dengan tipe teks yang berasal dari *database OAJIS* sendiri.

Pengambilan data dilakukan dengan mengunduh data secara langsung pada *database OAJIS* secara langsung dengan bantuan pengelola *website OAJIS*. Data yang digunakan merupakan data periodik dari tahun 2013 hingga 2017. Atribut data dari *file .xlsx* yang didapatkan dari repositori *website OAJIS* adalah nama *file*, judul, abstrak Indonesia, *author*, *author e-mail* dan *keyword* Indonesia. Kemudian isi dari *paper* diambil dari *file .pdf paper*.

Atribut data yang diambil dari *file .xlsx* adalah nama *file* dan *author* setiap dokumen untuk membuat hubungan antara *author* dan dokumen. Sedangkan atribut data yang diambil dari *file .pdf* dimulai dari abstrak dalam Bahasa Indonesia hingga kesimpulan. Hal ini dilakukan karena bagian *header* seperti judul, penulis, alamat penulis, *e-mail*, dan daftar pustaka di bagian akhir *paper* dianggap tidak relevan untuk dijadikan data teks yang akan diproses dengan *topic modeling*.

4.2 Metodologi Implementasi Penelitian

Sub bab ini menjelaskan metodologi yang digunakan dalam mengimplementasikan penelitian untuk mencapai tujuan penelitian. Tahapan implementasi ini dilakukan dengan penyesuaian pada komputasi secara otomatis menggunakan bahasa pemrograman *Python*. Terdapat lima tahapan dalam

melakukan implementasi penelitian, yaitu mempersiapkan data, pra-proses data, pemrosesan data atau *topic modeling* dengan menggunakan *Author-Topic Model*, validasi model menggunakan nilai *perlexity*, *topic coherence* dan analisis topik model.

4.2.1 Mempersiapkan Data

Tahap mempersiapkan data merupakan tahapan untuk mengolah data yang sudah dimiliki menjadi bentuk struktur data yang dibutuhkan oleh tahap proses data sehingga siap digunakan untuk melakukan tahap proses data dan analisis. Alur tahap persiapan data sesuai dengan Diagram 4.2.1.

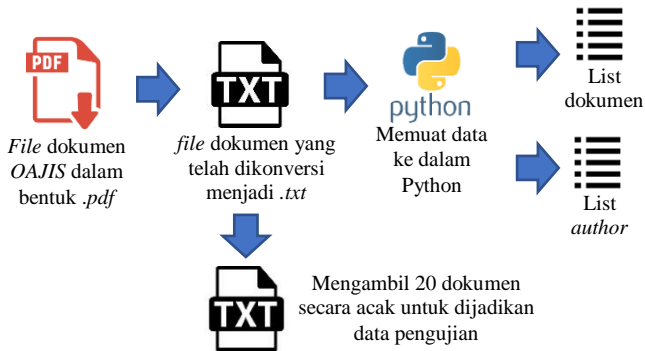


Diagram 4.2.1 Alur mempersiapkan data

Data mentah dalam format *.pdf* kemudian dikonversi menjadi *file .txt* menggunakan *converter online* untuk mempermudah proses pengolahan dengan bahasa pemrograman *Python*. Selanjutnya data yang sudah dalam format *.txt* kemudian dimuat ke dalam variabel pada *Python environment* dengan tipe list *array*. Data yang terdiri dari dokumen dipisah menjadi list dokumen dan list *author* pembuatnya.

Dari keseluruhan data yang berjumlah 431 dokumen, dilakukan pemisahan data sebagai data pelatihan dan data pengujian. Pemisahan data dilakukan dengan mengambil data pengujian secara acak yang kemudian disisihkan untuk diuji setelah model telah terbentuk. Syarat yang harus dipenuhi oleh data pengujian

adalah *author* dari dokumen pada data pengujian harus memiliki minimal satu dokumen yang ditulis pada data pelatihan. Hal ini dibutuhkan untuk melakukan pengecekan kemampuan prediksi model dalam mengetahui *author* asli dari dokumen pengujian berdasarkan topik yang dibentuk oleh model. Jumlah dokumen pengujian yaitu sebanyak 20 dokumen. Sedangkan jumlah dokumen pelatihan yaitu sebanyak 411 dokumen. Pembahasan selanjutnya merupakan tahapan yang dilakukan pada data pelatihan untuk membentuk model. Dokumen pengujian hanya akan digunakan ketika model-model dari hasil eksperimen telah terbentuk dan siap untuk diuji.

4.2.1.1 Membuat *list author*

Data dengan atribut nama *file*, judul, abstrak Indonesia, *author*, *author e-mail* dan *keyword* Indonesia dari *file .xlsx* diambil hanya nama *file* dan *author* saja untuk membuat *file list author* yang berisikan hubungan antara *author* dengan setiap dokumennya dan disimpan dalam *file .csv*. Tahap ini bertujuan untuk memberi tahu model hubungan yang terjadi antara *author* dan dokumen sehingga proses penentuan topik *author* dilakukan berdasarkan pemetaan hubungan antara *author* dan dokumen yang ditulisnya ini. Contoh isi *list author* sesuai dengan Gambar 4.2.1.

```
A.A. Gede Bagus Ariana,80
Abdul Karim,38
Abdul Muhaimin,259
Aboeryzal Ahmed Koesyairy,348
Achmad Affandi,343
Achmad Affandi,363
Achmad Solichin,30
Achmad Taufik K. Akbar,23
Achmad Yasid,110
Ada Udi Firmansyah,409
```

Gambar 4.2.1 *list author*

4.2.1.2 Membuat *list dokumen*

Data isi *paper* yang telah dikonversi dari *.pdf* menjadi *.txt* kemudian dimuat ke dalam variabel *array* di dalam *Python*

Environment. Nama *file* yang digunakan oleh setiap dokumen harus sesuai dengan peta hubungan antar *author* dan dokumen pada *list author* agar ketika dibentuk hubungan dalam *dictionary* pada model, setiap *author* dapat memanggil dokumen yang tepat sesuai dengan nama dokumen yang ditulis.

4.2.1.3 Membuat *author2doc*

Dari kedua *list* yang ada, dibuat hubungan *author2doc* yang memetakan tiap dokumen yang dimiliki oleh setiap *author* yang terdaftar ke dalam sebuah *dictionary*. Hubungan *author2doc* ini yang nantinya digunakan oleh model untuk membentuk distribusi topik pada *author* setelah model dan daftar topik terbentuk.

4.2.2 *Part-of-speech tag*

Pada tahapan ini dilakukan pemberian *tag* pada kelas kata yang ada pada data. Tahapan ini bertujuan untuk mengambil kelas kata kata benda karena bidang penelitian ilmiah seperti data yang digunakan lebih banyak tercerminkan dari kata benda yang tersebar dalam setiap dokumen. Pemberian *tag* ini dilakukan dengan menggunakan *library* Indonesia POS Tag yang dimiliki oleh Universitas Indonesia. Selanjutnya dilakukan penghapusan seluruh kata-kata yang tidak memiliki *tag* NN atau NNP, yaitu *tag* yang menggambarkan kata benda. Tahap pemberian *tag* pada data menggunakan POS Tag Indonesia sesuai dengan Diagram 4.2.2.

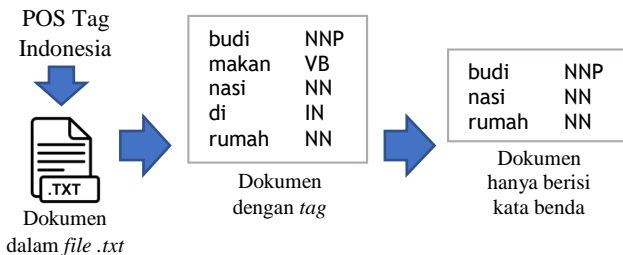


Diagram 4.2.2 Alur tahap pemberian tag pada data

4.2.3 Pra-Proses Data

Tahapan pra-proses data merupakan tahapan yang harus selalu dilakukan dalam penelitian penggalian data maupun *Natural Language Processing*. Tujuan dari pra-proses data adalah untuk menyeragamkan data dan membuang kata-kata yang tidak bermakna pada data. Pada tahapan ini dilakukan pengubahan data menjadi huruf kecil, *stemming* atau penghapusan imbuhan pada kata, *stopwords removal* atau penghapusan kata-kata yang tidak bermakna informasi, dan tokenisasi atau pemecahan kata-kata pada data menjadi token yang disimpan dalam *array*. Data yang melalui tahap pra-proses ini hanya data dokumen saja, karena *author* berisikan nama penulis. Alur pra-proses data dapat dilihat pada Diagram 4.2.3.

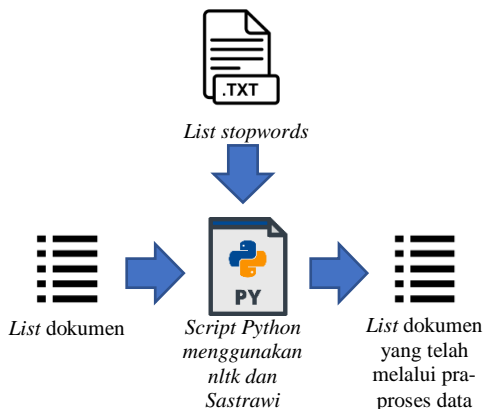


Diagram 4.2.3 Alur tahap pra-proses data

Pada tahap pengubahan data menjadi huruf kecil, tokenisasi, penghapusan karakter angka, dan penghapusan *stopwords* digunakan *library nltk*. Namun pendefinisian daftar *stopwords* yang digunakan pada tugas akhir ini mengacu pada susunan *stopword* yang telah disusun oleh Purwarianti, [19]. Sementara itu, untuk proses *stemming*, digunakan *library Sastrawi* di bawah lisensi MIT.

Setelah melalui pra-proses di atas, data dianggap sudah bersih dan siap untuk diproses pada tahap proses data.

4.2.4 Topic Modeling

Pada tahap ini mulai dilakukan pemrosesan data dari pembentukan *corpus* hingga penggunaan metode *Author-Topic Models* untuk melakukan klusterisasi pada dokumen.

4.2.4.1. Pembentukan *Dictionary* dan *Corpus*

Untuk melakukan *topic modeling* dengan metode *Author-Topic Model*, dibutuhkan pembentukan, *dictionary* dan *corpus* berdasarkan data yang telah dibersihkan pada tahap sebelumnya. *Dictionary* merupakan format data yang mengandung himpunan kata unik yang diberi indeks, untuk memudahkan proses menampilkan kata dalam model. Sedangkan *corpus* merupakan format data yang berbentuk dokumen *term-matrix*, untuk digunakan dalam pembentukan model. Tahapan ini dilakukan sesuai dengan alur pada Diagram 4.2.4.

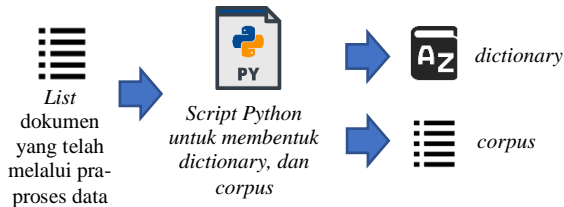


Diagram 4.2.4 Alur pembentukan *dictionary* dan *corpus*

4.2.4.2. Topic Modeling dengan *Author-Topic Model*

Pada tahap *topic modeling* dengan *Author-Topic Model*, langkah yang dilakukan adalah pembentukan model. Pada tugas akhir ini, *library Author-Topic Model* yang digunakan adalah *library author-topic model* dari gensim [18]. Setelah model terbentuk, nilai *perplexity* dan *topic coherence* dari setiap model dijadikan acuan untuk menentukan beberapa model terbaik. Alur proses tahapan ini sesuai dengan Diagram 4.2.5.

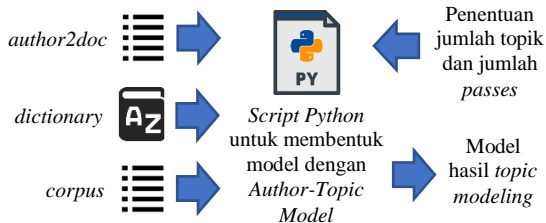


Diagram 4.2.5 Alur penentuan model

Membentuk model pada eksperimen dilakukan dengan menentukan jumlah topik yang menjadi kelompok klusterisasi kata dan penentuan jumlah *passes* proses hingga mencapai konvergen. Yang dimaksud jumlah *passes* dalam *topic modeling* merupakan jumlah iterasi model diulang dalam proses belajarnya. Proses eksperimen ini akan merujuk pada jumlah topik dan topik apa saja yang menyusun dokumen dan *author* yang dianalisa. Model dengan nilai *perplexity* terkecil dan tidak berubah-ubah akan dipilih menjadi model yang digunakan dan dianggap yang paling mendekati akurat.

4.2.5 Validasi Model

Tahapan ini bertujuan untuk memastikan model yang dibentuk dari proses *topic modeling* pada dokumen yang menghasilkan nilai probabilitas tertinggi, baik luaran berupa topik maupun kata-kata yang menyusun setiap topik. Beberapa hal yang diperhatikan dalam tahap validasi model adalah;

1. Jumlah *passes* yang tepat untuk membentuk model.
2. Jumlah topik yang sesuai berdasarkan nilai *perplexity* tiap model.
3. Jumlah topik yang sesuai berdasarkan nilai *topic coherence* tiap model.

4.3 Perancangan Pengujian Model

Perancangan pengujian model yang telah didapatkan dari hasil validasi topik bertujuan untuk mengetahui kemampuan masing-masing model dalam melakukan prediksi kemiripan dokumen pengujian dengan *author* yang menulis dokumen pelatihan.

Terdapat 2 tahapan pada proses pengujian ini, yaitu perhitungan rata-rata nilai probabilitas kata dalam dokumen untuk setiap topik dan kemiripan antara dokumen dengan *author* berdasarkan matriks probabilitas kata per topik menggunakan *Hellinger Distance*.

4.3.1 Probabilitas Kata dalam Dokumen per Topik

Pada tahap ini dilakukan perhitungan rata-rata nilai probabilitas kata dalam dokumen berdasarkan nilai probabilitas kata di setiap topik yang didapatkan dari hasil *topic modeling*. Perhitungan ini akan menghasilkan matriks yang dapat menggambarkan topik dan nilai probabilitas yang dimiliki setiap dokumen uji berdasarkan distribusi kata-katanya. Nilai probabilitas kemudian diurutkan dari yang paling tinggi, sehingga nilai probabilitas tertinggi menunjukkan topik yang paling dominan dalam dokumen tersebut. Alur perhitungan vektor probabilitas dokumen sesuai dengan Diagram 4.3.1

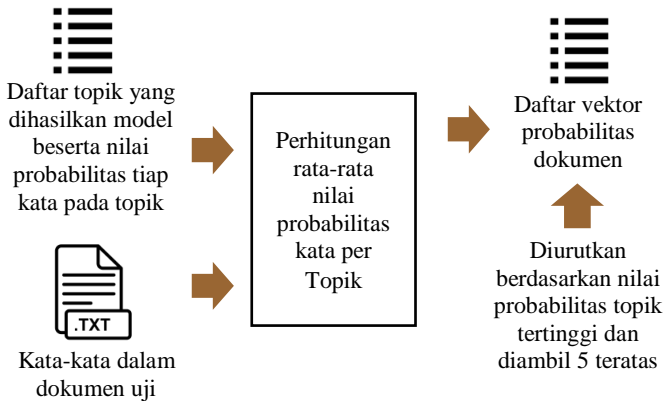


Diagram 4.3.1 Alur perhitungan vektor probabilitas dokumen

4.3.2 Perhitungan Kemiripan Dokumen dan *Author* berdasarkan Similaritas Vektor Probabilitas menggunakan *Hellinger Distance*

Setelah mengetahui nilai probabilitas dokumen untuk setiap topik yang menggambarkan dokumen tersebut, kemudian dibandingkan dengan nilai probabilitas dan topik yang menggambarkan *author* yang telah dihasilkan dari tahap pembentukan model. Perhitungan similaritas vektor probabilitas ini dilakukan menggunakan *Hellinger Distance*. *Hellinger Distance* ini merupakan perhitungan nilai kemiripan berdasarkan jarak pada vektor yang paling sering digunakan dalam perbandingan berdasarkan nilai probabilitas.

Halaman ini sengaja dikosongkan

BAB V IMPLEMENTASI

Bab ini menjelaskan implementasi dari perancangan yang telah dilakukan sesuai dengan metode pengembangan yang dibuat. Bagian implementasi ini akan menjelaskan mengenai lingkungan implementasi, pembuatan fitur-fitur aplikasi dalam bentuk kode, serta pengujian aplikasi.

5.1 Lingkungan Implementasi

Pengembangan aplikasi ini menggunakan komputer dengan spesifikasi sesuai Tabel 5.1.1.

Tabel 5.1.1 Spesifikasi Komputer

<i>Windows Based Operating Systems</i>	
Prosesor	Intel® Core™ i5-6200 CPU @2.30GHz
<i>Memory</i>	8 GB RAM
<i>Graphic Card</i>	AMD Radeon R7 M370 2GB
Sistem Operasi	<i>Windows 10 Home</i>
Arsitektur Sistem	<i>64-bit Operating System, x64-based processor</i>
<i>UNIX Based Operating Systems</i>	
Prosesor	Intel® Core™ i5 CPU @ 1.8 GHz
<i>Memory</i>	4 GB 1600 MHz DDR3
Sistem Operasi	OS X Yosemite 10.10.5

Model dikembangkan dengan menggunakan beberapa teknologi seperti editor, bahasa pemrograman, dan *library* yang disajikan dalam Tabel 5.1.2.

Tabel 5.1.2 Teknologi yang digunakan untuk mengembangkan model

Bahasa Pemrograman	<ul style="list-style-type: none"> • Python 3.6 • Perl 5.16.0 • Java JDK 9.0.1
Editor (IDE)	<ul style="list-style-type: none"> • <i>Notepad++</i> • <i>Visual Studio Code</i> • <i>Spyder</i> • <i>terminal mac bash v2.5.3</i>
Software/Tools	Microsoft Excel 2016
Library	<ul style="list-style-type: none"> • Anaconda 4.4 • Gensim 3.1 • Sastrawi 1.0.1 • Foma 0.9.18 • MorphInd v.1.4 • POS Tag Indonesia

5.2 Mempersiapkan Data

Sebelum memulai tahapan *topic modeling*, dilakukan persiapan data untuk diolah sehingga siap digunakan dalam pemrosesan. Terdapat tiga tahapan dalam mempersiapkan data, yaitu memberikan *tag* pada data untuk skenario data dengan *noun part-of-speech tag*, memuat data ke dalam *Python environment* dan membersihkan data.

5.2.1 Memberikan *tag* pada data untuk skenario data *topic modeling* dengan *noun part-of-speech tag*

Pada tahapan pemberian *tag* setiap kata yang terdapat dalam dokumen, proses *tagging* tidak dilakukan dalam bahasa pemrograman *Python*. Tahapan ini dilakukan dalam sistem operasi berbasis *UNIX* karena *library* POS Tag Indonesia yang digunakan hanya dapat dijalankan dalam sistem operasi berbasis *UNIX*.

Untuk dapat menjalankan *library* POS Tag Indonesia, terdapat beberapa *dependency* dan bahasa pemrograman yang dibutuhkan, yaitu *Java JDK 9.0.1*, *Perl 5.16.0*, *Foma 0.9.18*, dan *MorphInd v.1.4*. Setelah melakukan instalasi *Java JDK 9.0.1*, *Perl 5.16.0*, *Foma 0.9.18*, selanjutnya direktori *MorphInd v.1.4* dan POS Tag Indonesia harus tersusun seperti Gambar 5.2.1.

```
indonesian-tagger/
|-- tagger
|   |-- NER.pl
|   |-- testing.sh
|   `-- ...
`-- morphind
    |-- MorphInd.pl
    |-- bin-files
    |   |-- morphind.bin
    |   `-- ...
    `-- cache-files
        |-- 2432.tmp
        `-- ...
```

Gambar 5.2.1 Susunan direktori penggunaan POS Tag Indonesia

Direktori *tagger* merupakan direktori *library* POS Tag Indonesia dan direktori *morphind* adalah direktori *library* *MorphInd v.1.4* yang merupakan *dependency* dari POS Tag Indonesia.

Untuk menjalankan POS Tag Indonesia, digunakan *terminal mac bash v2.5.3*. Melalui terminal ini, masuk ke dalam direktori *tagger* dan memanggil *file tag.sh*. Selanjutnya, *file input* yang akan diproses menggunakan *tagger* dimasukkan ke dalam direktori *output* yang ada di dalam direktori *tagger* untuk dapat dipanggil oleh kode di dalam *file tag.sh*. *File input* ini harus diberi nama sesuai dengan format penamaan dari POS Tag Indonesia, yaitu *res-[id berkas]-input.txt*. Ketika menjalankan *tagger*, '[id berkas]' yang akan dipanggil dalam perintah menjalankan *file tag.sh*. Dalam implementasi penelitian ini, nama *file input* adalah *res-1-input.txt* hingga file ke-411. Kode yang digunakan untuk menjalankan *tagger* POS Tag Indonesia tertera pada kode 5.1.

```
$ cd tagger
$ ./tag.sh -file 1
```

Kode 5.1 Menjalankan POS Tag Indonesia

Hasil dari proses ini akan disimpan dalam direktori *outputs* dengan nama file *res-1-resolved.txt*. Hasil dari proses di atas dapat dilihat pada Gambar 5.2.2.

```
Sistem informasi      NN
yang SC
terintegrasi         NN
membutuhkan         VB
tata kelola          NN
dikarenakan         X
memberikan          VB
pengarahan          NN
dan CC
memastikan          VB
pencapaian          NN
kinerja NN
sesuai JJ
dengan SC,IN
tujuan NN
yang SC
diinginkan          VB
.
Z
```

Gambar 5.2.2 Hasil proses tagging dengan POS Tag Indonesia

Sesuai dengan tujuan penelitian, kata-kata yang akan digunakan hanyalah kata benda dengan label ‘NN’ atau ‘NNP’. Sehingga dari hasil *tagging* pada file *res-1-resolved.txt* kemudian diproses lagi untuk menghapus kata-kata yang tidak memiliki label ‘NN’ atau ‘NNP’. Hasil data yang hanya berisi kata benda kemudian disimpan ke dalam file baru dengan nama file *1.txt* untuk seluruh file hingga file ke-411. Kode yang digunakan untuk menghapus kata-kata yang tidak memiliki label ‘NN’ atau ‘NNP’ sesuai dengan kode 5.2.

```

datadir = '/Users/kirun/Anaconda3/data/outputs-try/'
for i in range(1,432):
    infile = datadir + 'res-' + str(i) + '-resolved.txt'
    outfile = datadir + 'clean/' + str(i) + '.txt'
    fin = open(infile, encoding='utf-8')
    fout = open(outfile, 'w+', encoding='utf-8')
    contents = fin.read()
    for line in contents.split('\n'):
        if 'NN' in line:
            fout.write(line.strip() + '\n')
    fin.close()
    fout.close()

```

Kode 5.2 Menghapus kata dengan tag selain 'NN' dan 'NNP'

Setelah dilakukan implementasi, diketahui bahwa terdapat beberapa frasa yang belum terdaftar dalam kamus POS Tag Indonesia, sehingga terdapat beberapa frasa yang berasal dari kata kerja tidak dikenali sebagai frasa, melainkan sebagai kata kerja. Berdasarkan hal ini, dilakukan penambahan frasa secara manual pada kamus POS Tag Indonesia, merujuk pada frasa-frasa yang sering muncul dalam seluruh dokumen.

5.2.2 Memuat Data

Tahapan memuat data dilakukan dengan menggunakan IDE Spyder. Terdapat dua data yang akan dimuat, yaitu data *list* dokumen dan data *list author* yang berisikan hubungan antara *author* dan dokumen yang ditulisnya. Data *list* dokumen memiliki format *.txt*, sedangkan data *list author* memiliki format *.csv*. Untuk dapat melakukan pembacaan data, dibutuhkan beberapa *library* yang harus dimasukkan terlebih dahulu, seperti terlihat pada kode 5.3.

```

import logging
logging.basicConfig(format='%(asctime)s :
%(levelname)s : %(message)s', level=logging.INFO)

import os, re, csv
from pprint import pprint

```

Kode 5.3 Import library logging, os, re, csv

Setelah memasukkan *library* yang dibutuhkan, kemudian data *list* dokumen dimuat dengan menggunakan kode 5.4. Setiap

dokumen *paper* tersimpan dalam satu *file .txt* sehingga seluruh *file* harus digabung ke dalam satu variabel. Data *list* dokumen ini disimpan dalam variabel ‘*docs*’ yang nantinya akan dilanjutkan untuk digunakan dalam tahapan pembersihan data dan pra-proses data.

```
# Folder containing all OAJIS papers.
data_oajist = '/Users/kirun/Anaconda3/data/oajis/txt-
training/' # set path

# Load data
docs = []
doc_ids = []
files = os.listdir(data_oajist)
for file in files:
    with open(data_oajist + file, errors='ignore',
encoding='utf-8') as fid:
        idsplit = file.split('.', 3)
        doc_id = idsplit[0]
        txt = idsplit[0] + ' ' + fid.read()
        txt = re.sub('\s', ' ', txt)
        doc_ids.append(doc_id)
        docs.append(txt)
```

Kode 5.4 Memuat *list* dokumen ke dalam *Python*

Pada tahapan ini, terdapat dua *list* dokumen yang dimuat, yaitu data tanpa *noun part-of-speech tag* dan data dengan *noun part-of-speech tag*. Kode yang digunakan sama, yang membedakan hanya *file* dokumen *.txt* yang dipanggil.

Setelah memuat data *list* dokumen, selanjutnya memuat data *list author* untuk nantinya dibuat hubungan *author2doc* untuk menghubungkan kedua data dalam model. Kode 5.5 merupakan kode yang digunakan untuk memuat data *list author*.

```

author2doc = dict()
i = 0
with
open('/Users/kirun/Anaconda3/data/oajis/authorfix-
tr.csv', errors='ignore', encoding='utf-8') as author:
    readCSV = csv.reader(author, delimiter=',')
    for contents in readCSV:
        authorname = contents[0]
        authorname = re.sub('\s', ' ', authorname)
        ids = contents[1:]
        if not author2doc.get(authorname):
            author2doc[authorname] = []
            author2doc[authorname].extend(ids)

```

Kode 5.5 Memuat *list author* ke dalam *Python*

Untuk membuat hubungan antara *author* dan *dokumen* dalam model, digunakan format *dictionary* dalam *python* seperti pada kode 5.6.

```

doc_id_dict = dict(zip(doc_ids, range(len(doc_ids))))
for a, a_doc_ids in author2doc.items():
    for i, doc_id in enumerate(a_doc_ids):
        author2doc[a][i] = doc_id_dict[doc_id]

```

Kode 5.6 Membuat hubungan *author2doc* antara *list author* dengan *list dokumen*

Setelah semua data dimuat dan hubungan antara *author* dan setiap dokumen sudah terbentuk, data *list dokumen* akan diproses lebih lanjut untuk tahapan pembersihan data dan pra-proses data. Sedangkan data *list author* tidak akan melalui kedua proses tersebut karena hanya berisikan daftar nama *author* dan dokumen yang ditulis oleh setiap *author*.

5.2.3 Membersihkan Data

Dalam melakukan tahapan pembersihan data, digunakan *library regular expression* seperti pada tahap sebelumnya. Data yang akan dibersihkan merupakan data *list* dokumen yang telah dimuat sebelumnya dan sudah tersimpan dalam variabel '*docs*'. Terdapat dua tahapan yang dilakukan pada pembersihan data ini, yaitu penghapusan karakter numerik dan simbol untuk hanya mengambil karakter *alpha* saja, dan penghapusan satu digit karakter. Hasil dari penghapusan karakter *non-alpha*

disimpan dalam variabel *'nonalpha'*, kemudian hasil pada variabel tersebut dilanjutkan diproses untuk menghapus satu digit karakter yang disimpan dalam variabel *'onedigit'*. Setelah setiap dokumen dalam *list* dokumen selesai dibersihkan, seluruh data dalam *list* disimpan dalam variabel *'cleaneddata'*. Implementasi proses pembersihan data sesuai dengan Kode 5.7.

```
# Data Cleaning
pattern = re.compile('[^a-zA-Z]')
cleaneddata = []
for line in docs:
    nonalpha = re.sub(pattern, ' ', line)
    onedigit = re.sub(r'\b\w{1,1}\b', '', nonalpha)
    cleaneddata.append(onedigit)
```

Kode 5.7 Membersihkan data

Variabel *'cleaneddata'* ini yang kemudian akan dipanggil dalam tahap pra-proses data.

5.3 Pra-proses Data

Tahap pra-proses data merupakan tahap yang dilakukan agar data yang digunakan sesuai dengan masukkan yang dibutuhkan untuk diproses dengan model. Terdapat lima tahapan yang dilakukan dalam pra-proses data, yaitu pendefinisian *stopword*, menjalankan *case folding*, *stemming*, tokenisasi, dan penghapusan *stopword*, pembuatan frasa berdasarkan kata-kata pada dokumen, pembuatan *dictionary* dari dokumen, dan konversi dokumen menjadi bentuk matriks dokumen atau *corpus*.

5.3.1 Pendefinisian *stopword*

Pendefinisian *stopword* dilakukan dengan menyimpan daftar *stopword* pada *file* dengan format *.csv* dan disimpan dengan satu kata per baris. Daftar *stopword* ini kemudian dimuat ke dalam variabel *list* dengan nama *'list_stopword'*. Karena *stopword* disimpan dalam *file .csv*, dibutuhkan *library csv* untuk digunakan memuat daftar *stopword*. Kode yang digunakan untuk mendefinisikan *stopword* terdapat pada Kode 5.8.


```

import csv

# Stopword List
list_stopword = []
with open('/Users/kirun/Anaconda3/stopword.csv') as
stopword:
    for line in stopword:
        list_stopword.append(line.strip())

```

Kode 5.8 Pendefinisian *stopword list*

5.3.2 *Case folding, stemming, tokenisasi, dan penghapusan stopwords*

Setelah dilakukan pendefinisian *stopword*, nantinya *stopword* ini akan digunakan untuk menghapus *stopword*. Sebelum melakukan penghapusan *stopword*, dilakukan proses *stemming* dan tokenisasi. Untuk dapat melakukan tahapan *stemming* dan tokenisasi, dilakukan pembuatan *stemmer* dan *tokenizer* dengan menggunakan *library Sastrawi.Stemmer* dengan modul *StemmerFactory* dan *nlk.tokenize* dengan modul *RegexpTokenizer*. Kode yang digunakan sesuai dengan Kode 5.9.

```

from nltk.tokenize import RegexpTokenizer
from Sastrawi.Stemmer.StemmerFactory import
StemmerFactory

# Create Stemmer
factory = StemmerFactory()
stemmer = factory.create_stemmer()

# Create Tokenizer
tokenlist_data = []
tokenizer = RegexpTokenizer(r'\w+')

```

Kode 5.9 Pembuatan *Stemmer* dan *Tokenizer*

Langkah selanjutnya adalah tahap *case folding, stemming, tokenisasi, dan penghapusan stopwords*. Dokumen yang telah dibersihkan dan tersimpan dalam variabel '*cleaneddata*' kemudian diproses untuk dilakukan *case folding*, diubah menjadi *lowercase* dengan tipe data *string*. Hasil *lowercase* ini disimpan dalam variabel '*lowerdocs*' dan dilanjutkan untuk proses *stemming*. Proses *stemming* dilakukan dengan *stemmer*

yang sudah dibuat sebelumnya, dan disimpan dalam variabel *'stemmeddocs'*. Data yang sudah ter-*stem* dalam variabel *'stemmeddocs'* kemudian ditokenisasi dengan *tokenizer* yang telah dibuat sebelumnya, sehingga kata-kata dalam dokumen dipecah dan disimpan dalam bentuk token per kata pada variabel *'tokens'*. Selanjutnya kata-kata pada daftar *stopword* yang sudah didefinisikan sebelumnya dihapus dari dokumen hingga hasil akhir token yang sudah bersih disimpan dalam variabel *'tokenlist_data'*. Data pada variabel *'tokenlist_data'* ini kemudian disimpan dalam variabel *'datastsw'* untuk kemudian diproses ke tahap selanjutnya. Kode yang digunakan dalam tahap implementasi ini tertera pada Kode 5.10.

```
# CaseFolding, Tokenizing, RemoveStopwords
for i in cleaneddata:
    lowerdocs = str(i.lower())
    stemmeddocs = stemmer.stem(lowerdocs) # stemming
    process
    tokens = tokenizer.tokenize(stemmeddocs)
    stopped_tokens = [i for i in tokens if not i in
list_stopword]
    tokenlist_data.append(stopped_tokens)

datastsw = tokenlist_data
```

Kode 5.10 *Case Folding, Stemming, Tokenisasi, dan penghapusan stopword dari dokumen*

5.3.3 Pembuatan frasa

Pada tahapan ini dilakukan penyusunan frasa kata yang paling sering muncul dalam seluruh dokumen. Untuk menyusun frasa, digunakan *library gensim.models* modul *Phrases*. Kemudian pembentukan *bigram* atau frasa yang terdiri dari dua kata dilakukan dengan memilih frasa yang muncul lebih dari 20 kali. Data yang digunakan merupakan data hasil pra-proses data yang sebelumnya sudah disimpan di dalam variabel *'datastsw'*. Setiap frasa yang terbentuk dari kata-kata yang ada pada dokumen, akan ditambahkan ke dalam masing-masing dokumen yang ada pada variabel *'datastsw'*. Kode implementasi penyusunan *bigram* tertera pada Kode 5.11.

```

from gensim.models import Phrases

# Create Phrases
bigram = Phrases(datastsw, min_count=20)
for idx in range(len(datastsw)):
    for token in bigram[datastsw[idx]]:
        if '_' in token:
            datastsw[idx].append(token)

```

Kode 5.11 Membuat frasa kata

5.3.4 Pembuatan *dictionary* dari dokumen

Setelah membuat frasa berdasarkan kata-kata pada dokumen, kemudian dokumen dibuat menjadi bentuk *dictionary* untuk memberikan indeks pada setiap token. Token yang disimpan di dalam *dictionary* merupakan *unique tokens* dari seluruh token yang ada. Pada tahap ini juga dilakukan penghapusan *unique tokens* yang muncul terlalu sering, yaitu lebih dari 90% jumlah dokumen, dan token yang terlalu jarang muncul, yaitu kurang dari 10 kali kemunculan. Implementasi tahap pembuatan *dictionary* sesuai dengan Kode 5.12.

```

from gensim.corpora import Dictionary

# Create Dictionary
dictionarystsw = Dictionary(datastsw)
max_freq = 0.9
min_wordcount = 10
dictionarystsw.filter_extremes(no_below=min_wordcount,
                               no_above=max_freq)

_stsw = dictionarystsw[0] # init dictionary.id2token

```

Kode 5.12 Membuat *dictionary unique tokens* dari dokumen

5.3.5 Pembuatan dokumen matriks atau *corpus*

Pada tahapan ini dilakukan pembuatan dokumen matriks atau *corpus*. *Corpus* dibangun berdasarkan *dictionary* yang sudah dibentuk terhadap token yang ada dalam dokumen variabel '*datastsw*'. Tahap pembuatan *corpus* dilakukan sesuai dengan Kode 5.13.

```
# Vectorize data
corpusstsw = [dictionarystsw.doc2bow(doc) for doc in
datastsw]
```

Kode 5.13 Pembuatan *corpus* berdasarkan *dictionary* dan dokumen

5.4 Pemodelan Topik dengan metode *Author-Topic Models*

Tahap pemodelan topik dengan metode *Author-Topic Models* merupakan tahapan membentuk model dan melakukan eksperimen model untuk mendapatkan model yang paling baik.

5.4.1 Pembentukan model topik

Tahap awal dari pemodelan topik dengan menggunakan metode *Author-Topic Models* adalah pembentukan model topik. Untuk membentuk model topik, digunakan *library* dari *gensim.models* modul *AuthorTopicModel*. Beberapa *input parameter* yang harus diperhatikan dalam membentuk model topik yaitu jumlah topik dan jumlah *passes*. Dari kedua nilai *input parameter* ini nantinya dilakukan eksperimen untuk mendapatkan model yang terbaik. Implementasi pembentukan model sesuai dengan Kode 5.14.

```
from gensim.models import AuthorTopicModel

model = AuthorTopicModel(corpus=corpusstsw,
num_topics=50, id2word=dictionarystsw.id2token,
author2doc=author2doc, passes=100, eval_every=10)
```

Kode 5.14 Pembentukan topik model dengan *Author-Topic Model*

5.4.2 Eksperimen pemodelan topik dengan *Author-Topic Model*

Pada tahapan eksperimen, pemodelan topik dengan metode *Author-Topic Models* ini dilakukan pembentukan model terbaik dengan melakukan eksperimen pada *input parameter*, yaitu *passes* dan jumlah topik. Berdasarkan hal ini, eksperimen akan dibedakan ke dalam dua tahap, yaitu penentuan jumlah *passes* dan penentuan jumlah topik.

5.4.2.1. Penentuan Jumlah *Passes*

Pada metode *Author-Topic Model*, *passes* merupakan istilah yang digunakan untuk menyebut iterasi atau perulangan yang dilakukan model untuk belajar. Dalam menjalankan eksperimen di bidang *topic modeling*, penentuan jumlah *passes* merupakan tahapan yang penting. Hal ini dilakukan untuk mendapatkan perulangan yang tepat hingga dapat dihasilkan model yang telah belajar dengan baik. Jika jumlah *passes* terlalu sedikit, model belum mencapai konvergen untuk mendistribusikan seluruh kata-kata yang mirip atau satu topik. Sedangkan jumlah *passes* yang berlebihan akan menyebabkan model menjadi *overfitting*.

Untuk mendapatkan jumlah *passes* yang tepat, eksperimen tidak hanya dilakukan sekali, namun dilakukan sebanyak lima kali dengan jumlah topik yang berbeda-beda. Hal ini dilakukan untuk mendapatkan jumlah *passes* yang stabil. Jumlah topik mula-mula yang digunakan untuk menentukan jumlah *passes* yaitu 20, 40, 60, 80, dan 100 topik. Berdasarkan eksperimen ini, nilai *perplexity* yang muncul akan dicatat untuk dianalisa tren nilai *perplexity* setiap *passes*.

Nilai *perplexity* setiap *passes* akan divisualisasi menggunakan *line chart* di *Microsoft Excel*. Berdasarkan visualisasi yang dilakukan, dapat dilihat nilai *perplexity passes* mulai stabil sejak *passes* ke berapa. *Passes* yang digunakan adalah *passes* dengan nilai *perplexity* paling awal setelah menunjukkan tren yang stabil. Untuk dapat menampilkan nilai *perplexity* di setiap *passes*, dibutuhkan *library logging* sehingga hasil yang keluar tercatat pada *log*. Implementasi tahapan eksperimen jumlah *passes* ini sesuai dengan Kode 5.15.

```

import logging
logging.basicConfig(format='%(asctime)s :
%(levelname)s : %(message)s', level=logging.INFO)

from gensim.models import AuthorTopicModel

model = AuthorTopicModel(corpus=corpusstsw,
num_topics=20, id2word=dictionarystsw.id2token,
author2doc=author2doc, passes=100, eval_every=10)

```

Kode 5.15 Penentuan Jumlah *Passes*

5.4.2.2. Penentuan Jumlah Topik

Setelah mendapatkan jumlah *passes* yang sesuai, dilanjutkan dengan eksperimen penentuan jumlah topik. Dalam melakukan eksperimen penentuan jumlah topik, dilakukan percobaan dengan jumlah topik mulai 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, hingga 150 topik. Hal ini dilakukan untuk mendapat nilai *perplexity* terendah. Nilai *perplexity* yang semakin rendah menunjukkan kemampuan probabilitas prediksi model yang semakin baik. Eksperimen jumlah topik tersebut diulang sebanyak 30 kali, kemudian dihitung rata-rata dan nilai standar deviasi dari 30 kali pengujian tersebut. Hal ini dilakukan karena jumlah populasi minimal untuk menghasilkan distribusi normal adalah 30.

Nilai rata-rata dan standar deviasi *perplexity* yang telah didapatkan kemudian divisualisasikan menggunakan *bar chart* di dalam *Microsoft Excel*. Nilai rata-rata dan standar deviasi paling rendah menunjukkan model yang lebih baik. Kode yang digunakan untuk mengimplementasikan tahap penentuan jumlah topik tertera pada Kode 5.16.

```

import logging
logging.basicConfig(format='%(asctime)s :
%(levelname)s : %(message)s', level=logging.INFO)

from gensim.models import AuthorTopicModel

model = AuthorTopicModel(corpus=corpusstsw,
num_topics=150, id2word=dictionarystsw.id2token,
author2doc=author2doc, passes=15, eval_every=10)

```

Kode 5.16 Penentuan Jumlah Topik

5.4.2.3. Menyimpan Model

Setelah mendapatkan model terbaik berdasarkan jumlah *passes* yang sesuai serta rata-rata dan standar deviasi dari nilai *perplexity* setiap model, kemudian model disimpan dalam format *.atmodel*. Kode yang digunakan untuk menyimpan model sesuai dengan Kode 5.17.

```
# Save the model
model150.save('/Users/kirun/Anaconda3/saved
model/model150.atmodel')
```

Kode 5.17 Menyimpan Model

5.5 Validasi Model Topik

Setelah mendapatkan model yang dianggap baik berdasarkan nilai *perplexity*, dilakukan validasi model topik dengan menggunakan nilai *topic coherence*. Terdapat dua tahapan yang dilakukan dalam melakukan validasi dengan menggunakan nilai dari *topic coherence*, yaitu mengetahui *Top Topics* dan menghitung rata-rata *coherence score* setiap model.

Sebelum melakukan validasi topik dengan perhitungan *top topics* dan rata-rata *coherence score*, dilakukan tahap *load* model dari model yang telah disimpan sebelumnya. Untuk melakukan *load* model, kode yang dipanggil sesuai dengan Kode 5.18.

```
# Load the model
model150 = AuthorTopicModel.load('/Users/kirun
/Anaconda3/saved model/model150.atmodel')
```

Kode 5.18 Load model

5.5.1 Memanggil hasil *Top Topics*

Pada tahapan ini, dilakukan pencarian topik terbaik dari setiap model dengan mengurutkan topik berdasarkan nilai *coherence score* setiap topik pada masing-masing model. Dalam metode *Author-Topic Model*, setiap topik yang muncul pada model disertai dengan nilai *coherence score* atau *topic difference*. Semakin tinggi nilai *coherence score* menunjukkan kelompok

kata dalam topik yang semakin mirip. Perintah yang digunakan untuk mendapatkan urutan topik terbaik pada model tertera pada Kode 5.19.

```
# Print Top Topics
top_topics50 = model150.top_topics(model150.corpus)
pprint(top_topics50)
```

Kode 5.19 Memanggil hasil *Top Topics*

5.5.2 Rata-rata *Coherence Score*

Setelah mengetahui urutan topik terbaik dalam setiap model, dilakukan perhitungan rata-rata *coherence score* dari setiap model. Hal ini dilakukan dengan menjumlahkan seluruh nilai *coherence score* masing-masing topik dan dibagi dengan jumlah topik yang ada. Implementasi perhitungan rata-rata *coherence score* model dapat dilihat pada Kode 5.20.

```
avg_tc50= sum([t[1] for t in top_topics50])/50
pprint(avg_tc50)
```

Kode 5.20 Menghitung rata-rata *coherence score*

5.6 Pengujian model dengan menggunakan nilai similaritas vektor probabilitas

Tahapan ini bertujuan untuk menguji kemampuan prediksi model dalam mengetahui *author* yang membuat dokumen tertentu dan kemiripan topik penelitian antar *author*. Hal ini dilakukan karena pembentukan model dibentuk berdasarkan probabilitas distribusi kata-kata dalam dokumen. Oleh karena itu, untuk mengetahui hubungan antara distribusi kata-kata pada topik dalam dokumen dengan distribusi topik terhadap masing-masing *author* dapat dilakukan dengan menghitung jarak keduanya menggunakan *vector similarity*.

Pada tahap ini terdapat empat hal yang dilakukan, yaitu memuat data pengujian, menghitung rata-rata probabilitas kata-kata dalam dokumen berdasarkan topik, menghitung nilai *vector similarity* antara dokumen dan *author* serta menghitung nilai *vector similarity* antar *author*.

5.6.1 Memuat data pengujian

Pada tahap ini dilakukan tahap memuat data pengujian ke dalam *Python*. Data pengujian telah dipisahkan sejak awal, yaitu sejumlah 20 data. Tahapan pra-proses yang sama dengan data pelatihan juga dilakukan pada data pengujian untuk mendapatkan data yang telah bersih dan berbentuk token. Perintah pada tahap ini sesuai dengan Kode 5.21.

```
data_oajistes = '/Users/kirun/Anaconda3/data/oajis
/txt-testing/'

docsTes = []
doc_idsTes = []
files = os.listdir(data_oajistes)
for filen in files:
    with open(data_oajistes + filen, errors='ignore',
encoding='utf-8') as fid:
        idsplit = filen.split('.', 3)
        doc_id = idsplit[0]
        txt = idsplit[0] + ' ' + fid.read()
        txt = re.sub('\s', ' ', txt)
        doc_idsTes.append(doc_id)
        docsTes.append(txt)
```

Kode 5.21 Memuat data pengujian

Hasil dari tahapan pembersihan data dan pra-proses data pada data pengujian kemudian disimpan dalam variabel '*dataTes*' untuk selanjutnya digunakan dalam tahap perhitungan rata-rata probabilitas kata dalam dokumen pada topik.

5.6.2 Perhitungan rata-rata probabilitas kata dalam dokumen untuk setiap topik

Setelah memuat data, dilakukan perhitungan rata-rata probabilitas kata dalam dokumen pada setiap topik. Nilai probabilitas setiap kata yang terdistribusi pada setiap topik dijumlahkan lalu dibagi sejumlah kata yang ada pada setiap dokumen. Nilai rata-rata probabilitas kata per topik disimpan dalam variabel dengan format *dictionary* karena topik disini bersifat unik. Hal ini dilakukan untuk memudahkan pemanggilan indeks setiap topik dengan masing-masing nilai probabilitasnya. Hasil dari perhitungan probabilitas kata setiap

dokumen pada masing-masing topik disimpan dalam *dictionary* 'topics_avg' dan dimasukkan ke dalam *array list*, yaitu variabel 'doc_vecs' untuk memudahkan proses pemanggilan nilai vektor probabilitas. Kode yang digunakan untuk menjalankan tahapan ini dapat dilihat pada Kode 5.22.

```

docAvg_list = []
docSum_list = []
for data in dataTes:
    topics_sum = dict()
    topics_avg = dict()
    for word in data:
        for (keys, values) in modelsw50.id2word.items():
            if values == word:
                probTopics = modelsw50.get_term_topics(keys)
                for (id_prob, val_prob) in probTopics:
                    if id_prob not in topics_sum:
                        topics_sum[id_prob] = val_prob
                    else:
                        topics_sum[id_prob] = topics_sum[id_prob] + val_prob
                topics_avg[id_prob] = topics_sum[id_prob] / len(data)
    docAvg_list.append(list(topics_avg.items()))
    docSum_list.append(list(topics_avg.items()))

```

Kode 5.22 Perhitungan rata-rata probabilitas kata per topik

Setelah mendapatkan nilai seluruh probabilitas kata per topik untuk setiap dokumen, selanjutnya nilai probabilitas diurutkan untuk mengetahui topik yang paling dominan dalam dokumen tersebut. Kemudian, hasil dari urutan tersebut diambil nilai lima teratas untuk kemudian dibandingkan dengan nilai probabilitas topik *author*. Implementasi pengurutan nilai probabilitas dan pengambilan lima nilai probabilitas teratas sesuai dengan Kode 5.23.

```

# Sort probabilities value
from operator import itemgetter
for data in docAvg_list:
    data.sort(key=itemgetter(1), reverse=True)

# get top 5 value
doc_vecs = []
for data in docAvg_list:
    doc_vec = data[:5]
    doc_vecs.append(doc_vec)

```

Kode 5.23 Mengurutkan dan mengambil nilai probabilitas 5 teratas

5.6.3 Menghitung nilai similaritas vektor probabilitas

Pada tahap ini dilakukan perhitungan nilai similaritas vektor antara dokumen dan *author* berdasarkan nilai probabilitas kata per topiknya. Untuk menghitung similaritas vektor probabilitas menggunakan *hellinger distance* dibutuhkan *library matutils*. Dalam implementasi tahapan ini digunakan Kode 5.24.

```

from gensim import matutils

# Make a list of all the author-topic distributions
author_vecs = [modelsw50.get_author_topics(author) for
author in modelsw50.id2author.values()]

# Get similarity between two vectors
def similarity(vec1, vec2):
    dist = matutils.hellinger(matutils.sparse2full(vec1,
modelsw50.num_topics), matutils.sparse2full(vec2,
modelsw50.num_topics))
    sim = 1.0 / (1.0 + dist)
    return sim

# get similarity of vector to all authors
def get_sims(vec):
    sims = [similarity(vec, vec2) for vec2 in author_vecs]
    return sims

```

Kode 5.24 Perhitungan nilai similaritas vektor probabilitas menggunakan *Hellinger Distance*

Selanjutnya, hasil perhitungan nilai similaritas vektor probabilitas akan ditampilkan dalam tabel yang berisikan nama *author*, nilai similaritas vektor probabilitas, dan jumlah dokumen yang dimiliki *author* dalam data pelatihan. *Author* yang ditampilkan dalam tabel akan diurutkan sesuai dengan kemiripannya dengan dokumen berdasarkan nilai similaritas vektor probabilitasnya. Untuk dapat menampilkan tabel agar mempermudah tahap analisa hasil pengujian, dibutuhkan *library pandas*. *Input* parameter yang dibutuhkan pada perhitungan ini adalah dokumen yang akan diuji dan jumlah minimal dokumen yang ditulis oleh *author*. Kode yang digunakan dalam tahap ini dapat dilihat pada Kode 5.25.

```

import pandas as pd

# Get table with similarities, author names, and size
def get_table(i, top_n=10, smallest_author=1):
    # Get similarities.
    sims = []
    sims = get_sims(doc_vecs[i])

    # Arrange author names, similarities, and author sizes
    in a list of tuples
    table = []
    for elem in enumerate(sims):
        author_name = modelsw50.id2author[elem[0]]
        sim = elem[1]
        author_size=len(modelsw50.author2doc[author_name])
        if author_size >= smallest_author:
            table.append((author_name, sim, author_size))

    # Make dataframe and retrieve top authors.
    df = pd.DataFrame(table, columns=['Author', 'Score',
        'Size'])
    df = df.sort_values('Score', ascending=False)[:top_n]

    return df

# Get similarities between document and authors
for n in range(0,20):
    table = get_table(n)
    pprint('Model Stemming 50 Topik: Dokumen ' + str(n+1))
    print(table)

```

Kode 5.25 Membuat tabel nilai similaritas vektor probabilitas antara dokumen dan *author*

Sedangkan untuk mengetahui topik yang terdistribusi dalam setiap *author* dapat digunakan Kode 5.26.

```

def show_author(name):
    print('\n%s' % name)
    print('Docs:', modelstsw50.author2doc[name])
    print('Topics:')
    pprint([(topic_labels[topic[0]], topic[1]) for topic
        in modelstsw50[name]])

show_author('authorname')

```

Kode 5.26 Memanggil distribusi topik *author*

BAB VI HASIL DAN PEMBAHASAN

Pada bab ini akan dijelaskan proses pengujian dan analisis terhadap hasil pengujian yang diperoleh dari proses implementasi yang telah dibahas pada bab sebelumnya.

6.1 Mempersiapkan Data

6.1.1 Memuat Data

Data publikasi yang didapatkan dari repositori *OAJIS* sejak tahun 2013 hingga 2017 adalah sebanyak 431 dokumen. Kemudian dari keseluruhan dokumen dibagi menjadi dua, yaitu 411 dokumen menjadi data pelatihan dan 20 dokumen menjadi data pengujian. Jumlah dokumen, kata, dan *author* yang digunakan dalam pengujian sesuai dengan Tabel 6.1.1.

Tabel 6.1.1 Jumlah data pelatihan

Jumlah Dokumen	Jumlah Kata	Jumlah <i>Author</i>
411	1117873	762

6.1.2 Pemberian *tag* pada data dengan POS Tag Indonesia untuk skenario data *topic modeling* dengan *noun part-of-speech tag*

Untuk mendapatkan hanya data kata benda pada setiap dokumen, dilakukan *tagging* menggunakan POS Tag Indonesia. Jumlah dokumen yang digunakan untuk data pelatihan adalah sejumlah 411 data, sedangkan jumlah kata yang terdapat dalam seluruh dokumen sebelum diberikan *tag* dan setelah diberikan *tag* yang hanya diambil kata bendanya saja dapat dilihat pada Tabel 6.1.2.

Tabel 6.1.2 Jumlah data *tag*

Sebelum diberi <i>tag</i>	Setelah diberi tak dan diambil hanya kata benda
1117873	464377

Data hasil *tagging* ini selanjutnya juga melalui tahap pra-proses yang sama dengan data asli, yaitu pembersihan data dan penghapusan *stopword*.

6.2 Pra-Proses Data

Tahap ini dilakukan pada ketiga skenario data. *Case folding*, tokenisasi, dan penghapusan *stopword* dilakukan pada seluruh skenario. Perbedaan dari ketiga skenario tersebut adalah skenario pertama: tidak dilakukan *stemming* pada kata-kata di dalam data, skenario kedua: dilakukan *stemming* pada kata-kata di dalam data, dan skenario ketiga: yaitu data yang telah ditandai atau diberi *tag* dan hanya diambil kata bendanya saja. Perubahan jumlah kata pada ketiga data setelah dilakukan tahap pra-proses dapat dilihat pada Tabel 6.2.1.

Tabel 6.2.1 Perubahan jumlah kata setelah pra-proses

Tahapan	Tanpa <i>Stemming</i>	<i>Stemming</i>	<i>Noun POS Tagged</i>
Jumlah asli	1117873	1117873	464377
Pembersihan data	1001777	1001777	456718
<i>Stemming</i>	-	1001777	-
Tokenisasi	1001777	1001777	456718
Penghapusan <i>stopword</i>	616307	606330	419889

Pada tahap pra-proses ini, jumlah kata sebelum dan setelah melalui tahap *stemming* adalah sama karena proses *stemming* tidak melakukan penghapusan kata maupun token. Namun terdapat beberapa kata yang setelah melalui tahap *stemming* berubah menjadi kata yang termasuk *stopword*, sehingga terdapat selisih jumlah kata antara yang sudah dan belum melalui tahap *stemming*.

6.3 Pembuatan *Dictionary* dari Dokumen

Pada pembuatan *dictionary*, seluruh token diambil secara unik sehingga menghasilkan *unique tokens* yang mewakili setiap kata yang ada pada keseluruhan token. Setiap *unique token* ini

memiliki *key* pada *dictionary* untuk memudahkan pembuatan *corpus*.

Setelah dibentuk *dictionary*, terdapat jumlah *unique tokens* pada masing-masing data sesuai dengan Tabel 6.3.1.

Tabel 6.3.1 Jumlah *unique tokens*

Tahapan	Tanpa <i>Stemming</i>	<i>Stemming</i>	<i>Noun POS Tagged</i>
Seluruh <i>unique tokens</i>	40298	34177	16245
min = 10 max = 0.9	4571	3497	1876

Jumlah akhir *unique tokens* inilah yang kemudian digunakan sebagai *dictionary* pada proses pemodelan dengan *Author-Topic Models* di tahap selanjutnya.

6.4 Pemodelan dengan *Author-Topic Model*

Pada pemodelan dengan *Author-Topic Model*, terdapat 3 skenario yang akan dilakukan dalam eksperimen berdasarkan 3 jenis data yang telah disiapkan sesuai dengan penjelasan sebelumnya. Untuk mendapatkan model terbaik berdasarkan 3 skenario yang ada, terlebih dahulu dilakukan penentuan jumlah *passes* atau iterasi dan penentuan jumlah topik yang membentuk model.

6.4.1 Penentuan Jumlah *Passes*

Penentuan jumlah *passes* dilakukan berdasarkan analisa terhadap kestabilan nilai *perplexity* setiap *passes*. Percobaan yang dilakukan dengan mengambil nilai *passes* yaitu 100 untuk dianalisa pergerakan nilai *perplexity*-nya. Percobaan ini dilakukan lima kali dengan jumlah topik yang berbeda, yaitu 20, 40, 60, dan 100 untuk memastikan kekonsistenan pergerakan nilai *passes* yang dihasilkan. Hasil nilai *perplexity* yang muncul dari hasil eksperimen kemudian dicatat dan ditampilkan dalam Diagram 6.4.1.

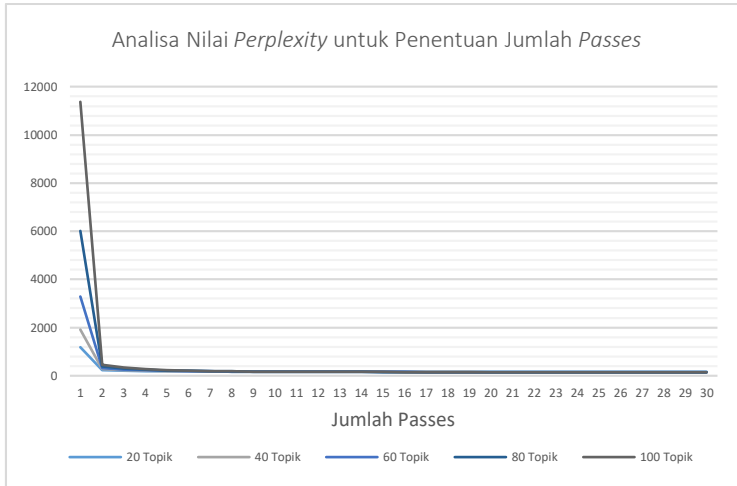


Diagram 6.4.1 Analisa nilai *perplexity* untuk menentukan jumlah *passes*

Dalam 100 *passes* awal yang ditentukan pada eksperimen, dari 30 *passes* saja sudah tampak kestabilan nilainya. Pada diagram di atas dapat dilihat bahwa pergerakan nilai *perplexity* telah stabil mulai *passes* ke-15. Hal ini juga tergambar dengan jelas oleh kestabilan nilai *passes* untuk kelima percobaan jumlah topik, sehingga dapat disimpulkan bahwa nilai *passes* yang digunakan pada tahap selanjutnya adalah 15.

6.4.2 Penentuan Jumlah Topik

Pada penentuan jumlah topik, juga dilakukan analisa berdasarkan nilai *perplexity*. Perbedaannya adalah pada penentuan jumlah *passes*, nilai *perplexity* yang diperhatikan adalah nilai *perplexity* setiap *passes* untuk dilihat kestabilannya. Sedangkan pada penentuan jumlah topik, yang diperhatikan adalah nilai *perplexity* terakhir yaitu pada *passes* ke-15, di mana nilai tersebut menunjukkan nilai *perplexity* model. Selain itu, dalam penentuan jumlah topik, eksperimen juga dilakukan dalam rentang jumlah topik yang lebih luas, yaitu dimulai dengan 10 topik hingga 150, dengan selisih setiap 10 topik pada masing-masing percobaannya. Eksperimen penentuan jumlah topik dilakukan pada ketiga skenario, yaitu skenario tanpa *stemming*, dengan *stemming*, dan *noun POS tagged*.

6.4.2.1. Tanpa Stemming

Hasil eksperimen nilai *perplexity* untuk penentuan jumlah topik pada skenario tanpa *stemming* dapat dilihat pada Lampiran 0.

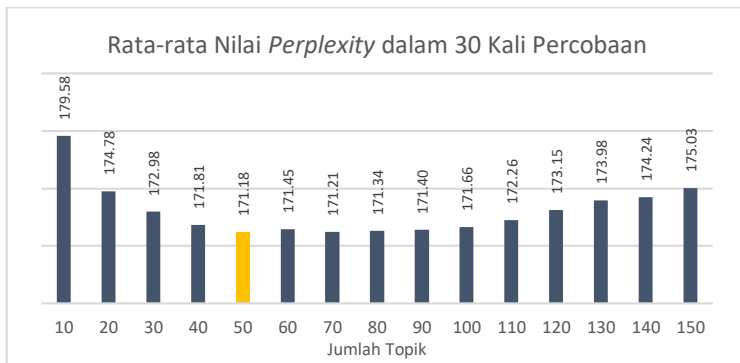


Diagram 6.4.2 Rata-rata nilai *perplexity* dalam 30 kali percobaan

Berdasarkan Diagram 6.4.2, rata-rata nilai *perplexity* paling rendah pada skenario tanpa *stemming* ini dihasilkan oleh model dengan jumlah topik 50 topik, yaitu 171.18. Pada jumlah topik yang rendah, tren nilai rata-rata sangat tinggi dan terus menurun hingga jumlah topik 50, kemudian kembali meningkat seiring dengan terus meningkatnya jumlah topik. Berdasarkan rata-rata nilai *perplexity* ini, dapat disimpulkan bahwa model dengan jumlah 50 topik merupakan model terbaik dalam skenario data tanpa *stemming*.

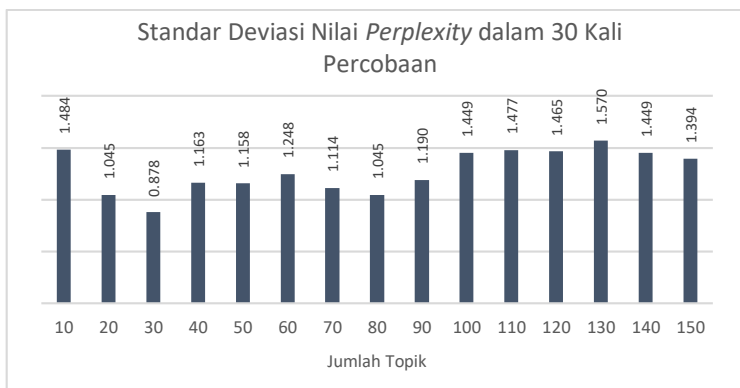


Diagram 6.4.3 Standar deviasi nilai *perplexity* dalam 30 kali percobaan

Sesuai dengan Diagram 6.4.3, dapat dilihat bahwa standar deviasi nilai *perplexity* model dengan jumlah topik 50 adalah 1.158, dimana nilai rata-rata standar deviasi nilai *perplexity* keseluruhan adalah 1.275. Jika keduanya dibandingkan, nilai 1.158 dapat dikatakan cukup stabil karena sedikit lebih rendah dari nilai rata-rata standar deviasi keseluruhan. Daftar 50 topik yang didapatkan dari model ini terlampir pada Lampiran B-1.

6.4.2.2. *Stemming*

Hasil eksperimen nilai *perplexity* untuk penentuan jumlah topik pada skenario dengan *stemming* dapat dilihat pada Lampiran 0.

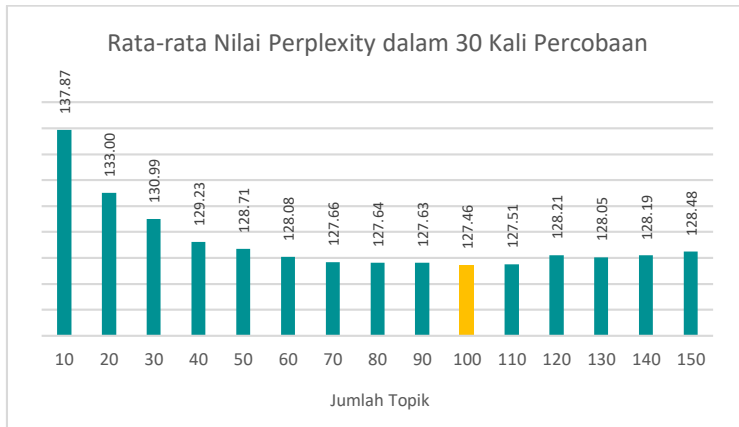


Diagram 6.4.4 Rata-rata nilai *perplexity* dalam 30 kali percobaan

Pada skenario data dengan *stemming*, dapat dilihat pada Diagram 6.4.4 bahwa model dengan rata-rata nilai *perplexity* paling rendah adalah model dengan jumlah topik 100, yaitu sebesar 127.46. Pada jumlah topik 10, rata-rata nilai *perplexity* sangat tinggi dan terus menurun hingga jumlah topik 100. Selanjutnya diketahui bahwa rata-rata nilai *perplexity* kembali meningkat hingga jumlah topik 150 walaupun tidak meningkat dengan signifikan, sehingga model dengan 100 topik merupakan model terbaik dalam skenario data dengan *stemming*.

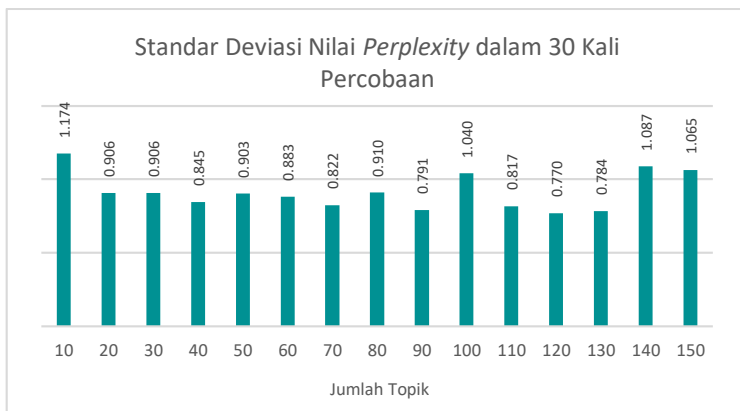


Diagram 6.4.5 Standar deviasi nilai perplexity dalam 30 kali percobaan

Selanjutnya berdasarkan Diagram 6.4.5, nilai standar deviasi nilai *perplexity* model dengan jumlah topik 100 adalah 1.040. Jika dibandingkan dengan rata-rata nilai standar deviasi keseluruhan, yaitu 0.914, nilai standar deviasi topik 100 sedikit lebih tinggi dari rata-rata nilai standar deviasi keseluruhan. Daftar 100 topik yang didapatkan dari model ini terlampir pada Lampiran B-2.

6.4.2.3. Noun POS Tagged

Hasil eksperimen nilai *perplexity* untuk penentuan jumlah topik pada skenario *noun POS tagged* dapat dilihat pada Lampiran 0.

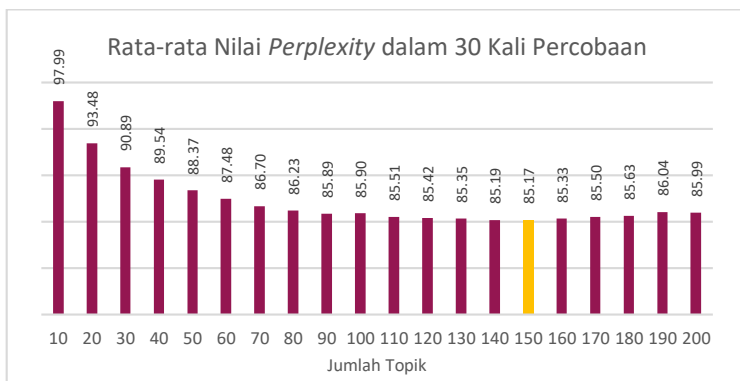


Diagram 6.4.6 Rata-rata nilai perplexity dalam 30 kali percobaan

Kemudian diketahui dari Diagram 6.4.6 bahwa pada skenario data yang hanya berisi kata benda dari keseluruhan dokumen berdasarkan hasil proses *POS tagging*, model dengan rata-rata nilai *perplexity* terendah adalah model dengan jumlah topik 150. Seiring dengan bertambahnya jumlah topik, rata-rata nilai *perplexity* terus menurun hingga mulai stabil pada jumlah topik 110 dan sampai pada nilai terendah yaitu 85.17, pada model dengan jumlah topik 150. Ketika eksperimen dilanjutkan hingga 200 topik, terlihat bahwa rata-rata nilai *perplexity* kembali meningkat. Berdasarkan hal ini, model dengan jumlah topik 150 merupakan model terbaik dalam skenario data dengan *noun POS tagged*.

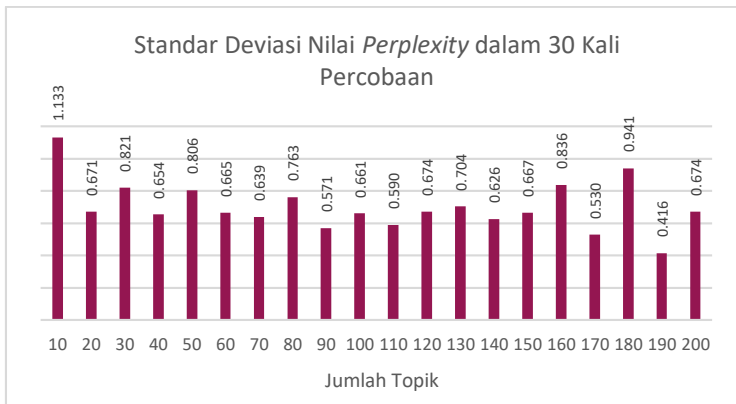


Diagram 6.4.7 Standar deviasi nilai perplexity dalam 30 kali percobaan

Jika dilihat nilai standar deviasinya, berdasarkan Diagram 6.4.7, model dengan 150 topik memiliki nilai standar deviasi sebesar 0.667. Nilai ini lebih rendah dari rata-rata nilai standar deviasi model keseluruhan, yaitu 0.702, sehingga nilai *perplexity* model 150 dapat dikatakan cukup stabil. Daftar 150 topik yang didapatkan dari model ini terlampir pada Lampiran B-3.

Berdasarkan ketiga eksperimen ini, didapatkan model terbaik dari masing-masing skenario, yaitu model dengan 50 topik pada skenario data tanpa *stemming*, model dengan 100 topik pada skenario data dengan *stemming*, dan model dengan 150 topik pada skenario data dengan *noun POS tagged*. Rata-rata dan

standar deviasi nilai *perplexity* hasil eksperimen yang didapatkan dari setiap model dapat dilihat pada Tabel 6.4.1.

Tabel 6.4.1 Rata-rata dan standar deviasi nilai *perplexity* model

	Tanpa <i>Stemming</i>	<i>Stemming</i>	<i>Noun POS</i> <i>tagged</i>
Jumlah Topik	50	100	150
Rata-rata nilai <i>perplexity</i>	171.18	127.46	85.17
Std. Deviasi nilai <i>perplexity</i>	1.158	1.040	0.667

6.5 Validasi Model Topik

Pada validasi model topik dilakukan perhitungan 10 topik terbaik yang dihasilkan oleh setiap model beserta nilai *coherence score* berdasarkan tinjauan pustaka mengenai *topic coherence*. Karena pada tahap sebelumnya didapatkan jumlah topik yang berbeda dari setiap skenario, setiap jumlah topik pada masing-masing skenario dibandingkan dengan model yang memiliki jumlah topik sama pada dua skenario lainnya. Hal ini dilakukan agar tahap perbandingan lebih seimbang dengan membandingkan model dengan parameter yang sama.

6.5.1 *Top Topics*

Pada tahap ini dilakukan pengurutan 10 topik teratas dari masing-masing model berdasarkan nilai *topic coherence*. Semakin tinggi nilai *topic coherence*, menjelaskan bahwa kata-kata dalam topik tersebut semakin baik dan mendekati kualitas yang sesuai dengan interpretasi manusia [14]. Hasil pengurutan 10 topik terbaik yang dihasilkan oleh setiap model dapat dilihat pada Lampiran C.

6.5.2 Nilai *Topic Coherence*

Berdasarkan nilai *topic coherence* yang didapatkan dari pemanggilan fungsi *top topics*, dapat dihitung rata-rata nilai *topic coherence* yang dimiliki oleh setiap model. Dalam menganalisis nilai *topic coherence*, semakin tinggi nilai menunjukkan hasil yang semakin baik. Rata-rata nilai *topic coherence* dari setiap model dapat dilihat pada Diagram 6.5.1.

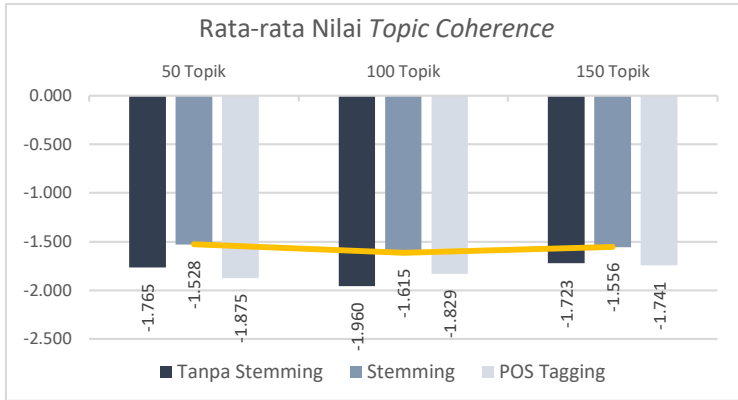


Diagram 6.5.1 Rata-rata nilai *topic coherence*

Pada Diagram 6.5.1, diketahui bahwa model dengan skenario *stemming* memiliki nilai rata-rata nilai *topic coherence* yang paling tinggi untuk setiap jumlah topik, yaitu -1.528 untuk model dengan jumlah topik 50, -1.615 untuk model dengan jumlah topik 100, dan -1.556 untuk model dengan jumlah topik 150. Untuk melihat rentang nilai *topic coherence* dari setiap topik dapat dilakukan visualisasi nilai standar deviasi, seperti pada Diagram 6.5.2.

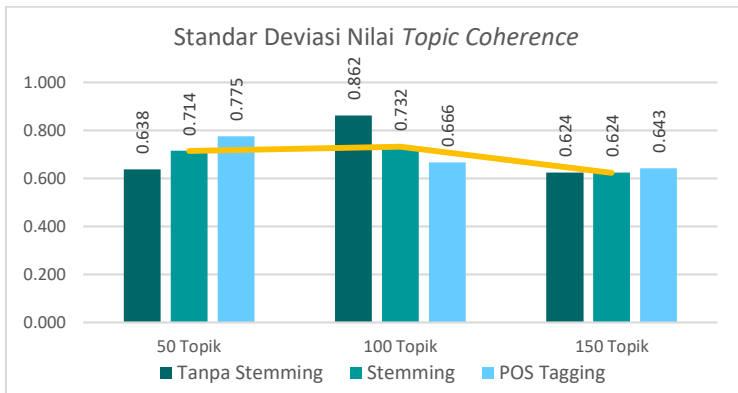


Diagram 6.5.2 Standar deviasi nilai *topic coherence*

Berdasarkan standar deviasi pada Diagram 6.5.2, diketahui bahwa skenario *stemming* memiliki nilai yang stabil, yaitu 0.7, jika dibandingkan dengan 2 skenario lainnya yang lebih

fluktuatif. Dengan nilai standar deviasi yang rendah, menjelaskan bahwa nilai *topic coherence* skenario *stemming* memiliki rentang yang tidak terlalu lebar, sehingga dapat mendukung nilai rata-rata yang paling tinggi seperti pada Diagram 6.5.1.

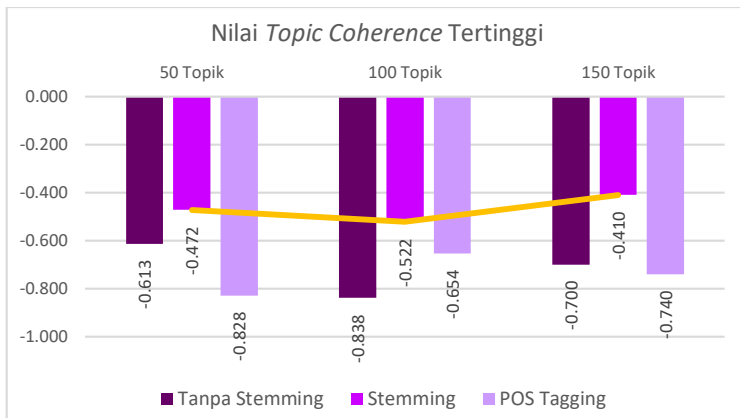


Diagram 6.5.3 Nilai *topic coherence* tertinggi

Pada Diagram 6.5.3 juga diketahui bahwa skenario *stemming* memiliki nilai *topic coherence* maksimal yang paling tinggi, yaitu sekitar -0.4. Hal ini menjelaskan nilai rata-rata yang menunjukkan bahwa model pada skenario *stemming* memiliki kualitas topik yang baik jika dikaitkan dengan interpretasi manusia.

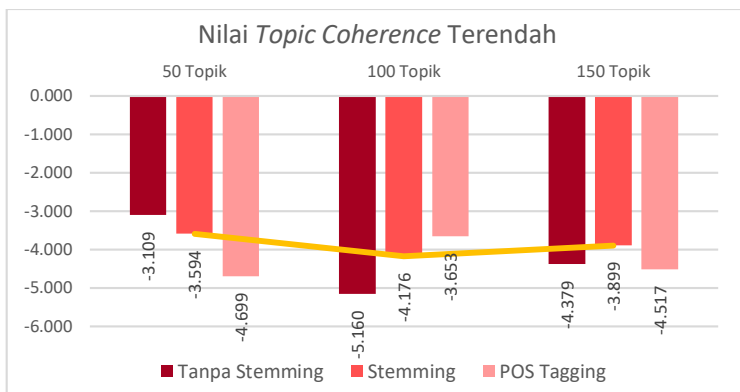


Diagram 6.5.4 Nilai *topic coherence* terendah

Kemudian pada Diagram 6.5.4 terlihat bahwa skenario *stemming* juga memiliki nilai *topic coherence* terendah yang tidak terlalu buruk dibandingkan dengan 2 skenario lainnya. Nilai ini pun cenderung stabil di angka -3.5.

Berdasarkan diagram yang telah dijelaskan di atas, dapat disimpulkan bahwa model dengan skenario *stemming* memiliki nilai *topic coherence* yang lebih baik dibandingkan dengan model pada 2 skenario lainnya. Kemudian pada Diagram 6.5.1 juga diketahui bahwa pada skenario *stemming*, model dengan nilai *topic coherence* tertinggi adalah model dengan jumlah topik 50.

Jika dikaitkan dengan rata-rata nilai *perplexity* pada tahap penentuan jumlah topik, memang jumlah topik yang memiliki rata-rata nilai *perplexity* paling rendah adalah model dengan jumlah topik 100. Namun ketika diperhatikan kembali pada Diagram 6.4.4, rata-rata nilai *perplexity* sejak topik berjumlah 50 mulai stabil hingga topik berjumlah 100. Model dengan jumlah topik 50 memiliki rata-rata nilai *perplexity* 128.71, dimana nilai ini hanya memiliki selisih sebesar 1.25 dengan rata-rata nilai *perplexity* model terendah dengan jumlah topik 100, yaitu 127.46.

Dari tahap validasi model berdasarkan nilai *perplexity* yang menjelaskan tingkat ketidakmiripan antar topik dan nilai *topic coherence* yang menjelaskan kualitas topik berdasarkan interpretasi manusia, didapatkan model *stemming* dengan jumlah topik 50 dengan nilai *perplexity* yang rendah, yaitu itu 128.71 dan nilai *topic coherence* tertinggi yaitu -1.528. Kedua nilai ini menjelaskan bahwa model *stemming* dengan jumlah topik 50 memiliki topik-topik yang lebih tidak mirip dibandingkan dengan model lain dan lebih sesuai dengan *human judgement* atau penilaian manusia.

6.6 Pengujian Model

Pengujian model dilakukan pada 20 data pengujian yang telah dipisahkan sebelumnya. Pemisahan dilakukan dengan memilih dokumen secara acak, dengan syarat *author* pada data pengujian

telah terdaftar sebagai *author* dokumen pada data pelatihan. Daftar 20 data yang digunakan pada tahap pengujian model dapat dilihat pada Lampiran D-1.

Pada 9 model yang telah didapatkan, masing-masing diuji 3 kali dengan nilai *minimal author size* 1, 3 dan 5. *Author size* merupakan jumlah dokumen yang ditulis oleh setiap *author* dalam data pelatihan. Sedangkan nilai *minimal author size* pada saat menampilkan tabel merupakan jumlah minimal dokumen yang ditulis oleh *author* ketika menampilkan daftar *author* yang mirip dengan dokumen pengujian dalam tabel similaritas vektor probabilitas.

Contoh tabel hasil pengujian dokumen dengan similaritas vektor probabilitas pada salah satu model dapat dilihat pada Gambar 6.6.1.

Model Stemming 50 Topik: Dokumen 3				
	Author	Score	Size	
20	Hanim Maria Astuti	0.599362	5	
37	Stanley Karouw	0.594896	3	
38	Titus Kristanto	0.594796	4	
30	Raras Tyasnurita	0.594454	3	
28	Oleh Soleh	0.594409	3	
17	Feby Artwodini Muqtadiroh	0.594351	8	
18	Fransiskus Adikara	0.594349	4	
29	Radityo Prasetyanto Wibowo	0.593639	3	
35	Safrian Aswati	0.592554	8	
32	Rida Indah Fariani	0.591704	3	

Gambar 6.6.1 Contoh hasil pengujian similaritas vektor probabilitas

Model yang diuji pada Gambar 6.6.1 merupakan model dengan skenario *Stemming*, jumlah topik 50, dan *minimal author size* yang digunakan adalah 3. Dokumen uji pada contoh Gambar 6.1 merupakan dokumen 3, dengan judul “Perencanaan Strategis Teknologi Informasi pada Sektor Publik menggunakan Kerangka *The Open Group Architecture Framework (TOGAF)*”. Penulis dari dokumen ini adalah Mario Glendi Kasenda, Eko Nugroho, dan Selo Sulistyو. Pada hasil pengujian, penulis asli dari dokumen tidak muncul, karena penulis hanya memiliki 1 dokumen yang terdaftar dalam data pelatihan, sedangkan *minimal author size* yang diberikan adalah 3, sehingga tabel hanya akan menampilkan 10 *author* paling

mirip dengan *minimal author size* 3. Hasil pengujian dokumen terhadap masing-masing model terlampir pada LAMPIRAN D.

Setelah dilakukan pengujian, hasil 10 *author* paling mirip berdasarkan nilai similaritas vektor probabilitas yang dihasilkan oleh pengujian dari setiap model disimpan dan dihitung jumlah *author* asli yang dapat diprediksi oleh masing-masing model. Kemampuan model melakukan prediksi *author* dokumen-dokumen pengujian dilakukan dengan melihat apakah *author* dari dokumen pengujian muncul dalam daftar 10 *author* termirip tersebut. Rumus perhitungan *recall author* dokumen pengujian sesuai pada persamaan (1).

$$\text{Recall author} = \frac{\text{Jumlah dokumen yang diprediksi dengan benar}}{\text{Jumlah dokumen pengujian}} \times 100\% \quad (1)$$

Untuk mengetahui *recall* atau kemampuan prediksi model, jumlah dokumen yang *author*-nya dapat diprediksi dengan benar akan dibagi dengan jumlah dokumen pengujian, yaitu 20. Nilai ini kemudian akan dikali dengan 100% untuk mendapatkan nilai persentasenya. Hasil perhitungan kemampuan prediksi model untuk setiap pengujian yang dilakukan dapat dilihat pada Tabel 6.6.1.

Tabel 6.6.1 Recall author dokumen pengujian

<i>Min author size</i>	<i>Tanpa Stemming</i>			<i>Stemming</i>			<i>Noun POS Tagged</i>		
	50	100	150	50	100	150	50	100	150
1	0%	0%	0%	5%	0%	0%	5%	0%	0%
3	15%	10%	25%	15%	25%	10%	15%	15%	20%
5	20%	40%	30%	25%	25%	25%	25%	30%	25%

Secara umum, berdasarkan hasil pengujian yang telah dilakukan, dapat dilihat pada LAMPIRAN D bahwa pada jumlah topik dan nilai *minimal author size* yang semakin besar, terjadi bias ketika menampilkan 10 *author* yang memiliki kemiripan paling tinggi pada dokumen. Hal ini dibuktikan dengan 10 *author* yang ditampilkan untuk setiap dokumen sebagian besar memiliki urutan yang sama, atau memiliki urutan yang berbeda dengan nama-nama yang sama. Oleh

karena itu, hasil pengujian dengan *minimal author size 5* yang menghasilkan daftar *author* yang serupa diberikan tanda warna merah.

6.6.1 Analisis Hasil berdasarkan Skenario Data

Pada analisis ini dibandingkan hasil kemampuan prediksi model berdasarkan skenario data yang telah ditentukan sejak awal, yaitu data tanpa *stemming*, dengan *stemming*, dan *noun POS tagged*. Setelah dilakukan pengujian pada setiap model, didapatkan hasil kemampuan prediksi model sesuai dengan Tabel 6.6.1. Pada hasil pengujian ini, diketahui bahwa model dengan skenario *stemming* memiliki nilai kemampuan prediksi yang lebih tinggi dibandingkan dua skenario lainnya. Hal ini dibuktikan dengan kemampuan model dalam memprediksi *author* dokumen dengan batasan nilai *minimal author size 1*. Model dengan *stemming* berhasil menebak 1 *author* dari 20 dokumen yang ada dengan benar. Walaupun skenario dengan *noun POS Tagged* juga berhasil menebak 1 *author* dari 20 dokumen dengan batasan nilai *minimal author size 1* dengan benar, tetapi skenario ini memiliki bias yang lebih tinggi pada hasil pengujiannya, dan nilai persentase kemampuan prediksi model yang lebih rendah dibandingkan dengan skenario dengan *stemming*.

Jika nilai persentase kemampuan prediksi model pada masing-masing skenario data dijumlahkan, diketahui bahwa skenario dengan *stemming* dapat memprediksi lebih banyak dokumen dibanding dua skenario lainnya, yaitu sekitar 14 dokumen yang dapat diprediksi dengan benar dari 9 kali pengujian yang dilakukan.

6.6.2 Analisis Hasil berdasarkan Minimal Author Size

Pada analisis ini dibandingkan hasil kemampuan prediksi model berdasarkan *minimal author size* yang ditentukan pada setiap kali pengujian. Hal ini dilakukan untuk melihat performa model dalam melakukan prediksi *author* berdasarkan batasan *minimal author size* yang diberikan. Semakin model dapat menebak *author* dokumen pada nilai *minimal author size* yang kecil,

berarti kemampuan prediksi model semakin baik. Karena semakin sedikit nilai *minimal author size*, berarti model harus bisa memprediksi *author* dengan lebih spesifik.

Pada Tabel 6.7 dapat diketahui bahwa pada *minimal author size* 1, hanya 2 dari 9 pengujian yang dapat memprediksi dengan benar. Sedangkan pada *minimal author size* 3 dan 5, 9 pengujian yang dilakukan pada masing-masing batasan menghasilkan hasil pengujian yang berbeda-beda. Pengujian dengan *minimal author size* 3 memiliki persentase kemampuan prediksi yang lebih tinggi dari pengujian dengan *minimal author size* 5. Pengujian dengan *minimal author size* 5 justru menyebabkan hasil prediksi mengalami bias atau tidak valid. Hal ini dikarenakan oleh jumlah dokumen terbanyak yang dibuat oleh *author* pada data pelatihan adalah 12. Sehingga ketika akan ditampilkan 10 *author* yang paling mirip dengan dokumen yang diuji, jumlah *author* dengan batasan 5 sudah terlalu sedikit. Sedangkan *author* dengan jumlah dokumen 1 jauh lebih banyak dibandingkan dengan *author* dengan jumlah dokumen lebih dari 3. Untuk mempermudah, jumlah dokumen yang ditulis *author* dapat dilihat pada Diagram 6.6.1.

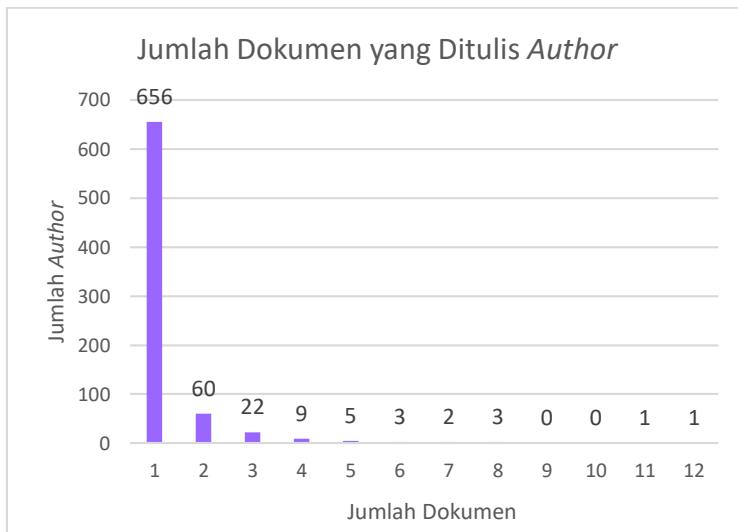


Diagram 6.6.1 Histogram jumlah dokumen yang ditulis oleh *author*

Pada Diagram 6.6.1 terlihat bahwa jumlah *author* yang hanya menulis 1 dokumen sangat banyak hingga mencapai 656 sedangkan sisanya menulis 2 sampai 3 dokumen, dan hanya 1 *author* yang menulis dokumen terbanyak, yaitu 12. Jumlah dokumen terbanyak yang ditulis oleh *author* adalah 12, sedangkan jumlah dokumen tersedikitnya adalah 1, dengan rata-rata 1.288 dan nilai standar deviasi 0.999. Berdasarkan nilai rata-rata, terlihat jelas bahwa jauh lebih banyak jumlah *author* dengan jumlah dokumen yang ditulis sama dengan 1 dibanding jumlah lainnya. Nilai standar deviasi yang cukup tinggi juga menunjukkan bahwa rentang jumlah dokumen yang ditulis oleh *author* terlalu lebar, yaitu dari 1 hingga 12 dengan jumlah yang tidak seimbang.

Jika dikaitkan dengan batasan pengujian *minimal author size*, nilai 5 jelas membatasi hasil *query author* terlalu banyak, karena 5 sudah hampir setengah dari 12 dan jumlah *author* dengan *minimal author size* 5 pun tidak banyak. Sedangkan pengujian menggunakan *minimal author size* 1 terlalu meluaskan rentang hasil *query author*, karena jumlah *author* dengan nilai *minimal author* 1 sangatlah banyak, mendekati jumlah total *author*. Oleh karena itu nilai *minimal author size* yang sesuai adalah 3, sesuai pada Tabel 6.7 yang menghasilkan persentase kemampuan prediksi model paling tinggi. Pada *author* dengan jumlah dokumen 3 ke atas juga model dapat belajar lebih baik karena menyebabkan dokumen yang dipelajari tidak terlampaui spesifik namun tidak terlampaui umum juga.

Berdasarkan ketiga perilaku pengujian yang dilakukan pada model, diketahui model terbaik untuk dapat memprediksi similaritas atau kemiripan dokumen pengujian dan *author* yang terdaftar dengan data publikasi berbahasa Indonesia yang dimiliki oleh repositori *OAJIS* adalah model dengan skenario *stemming* dan nilai *minimal author size* yang tidak terlalu kecil maupun besar, yaitu 3. Namun dalam proses pengerjaannya, banyak faktor yang harus diperhatikan ketika melakukan *author-topic modeling* menggunakan metode ini, seperti distribusi jumlah dokumen yang ditulis oleh *author* dan

kekonsistenan *author* dalam menulis penelitian atau publikasinya.

Ketika standar pengukurannya diturunkan dari kemampuan prediksi model dalam mengetahui *author* yang menulis dokumen pengujian tersebut menjadi kemampuan prediksi model dalam mengetahui *author* yang menulis dokumen yang mirip dengan dokumen pengujian, persentase tentu akan meningkat. Hal ini terjadi di beberapa kasus, contohnya adalah pada dokumen pengujian 3 di model *stemming* 50 topik dengan *minimal author size* 3. Dokumen ini memiliki judul “Perencanaan Strategis Teknologi Informasi pada Sektor Publik menggunakan Kerangka *The Open Group Architecture Framework (TOGAF)*”. Penulis dari dokumen ini adalah Mario Glendi Kasenda, Eko Nugroho, dan Selo Sulistyio. Penulis hanya memiliki 1 dokumen yang terdaftar dalam data pelatihan, sehingga dapat dilihat pada Gambar 6.6.2, bahwa penulis asli dari dokumen tidak muncul.

Model Stemming 50 Topik: Dokumen 3

	Author	Score	Size
20	<u>Hanim Maria Astuti</u>	0.599362	5
37	<u>Stanley Karouw</u>	0.594896	3
38	<u>Titus Kristanto</u>	0.594796	4
30	<u>Raras Tyasnurita</u>	0.594454	3
28	<u>Oleh Soleh</u>	0.594409	3
17	<u>Feby Artwodini Muqtadiroh</u>	0.594351	8
18	<u>Fransiskus Adikara</u>	0.594349	4
29	<u>Radityo Prasetianto Wibowo</u>	0.593639	3
35	<u>Safrian Aswati</u>	0.592554	8
32	<u>Rida Indah Fariani</u>	0.591704	3

Gambar 6.6.2 Hasil pengujian dokumen 3

Berdasarkan kasus yang terjadi ini, dapat dilakukan pengecekan vektor probabilitas dokumen 3 untuk mengetahui topik teratas yang terdistribusi pada dokumen ini. Berdasarkan vektor probabilitasnya, dokumen 3 memiliki distribusi topik sesuai dengan

Tabel 6.6.2.

Tabel 6.6.2 Distribusi topik dokumen 3

Dokumen 3	
Topik 40	0.00342
Topik 43	0.00145
Topik 42	0.00140
Topik 7	0.00117
Topik 4	0.00065

Kemudian dilakukan pengecekan topik yang terdistribusi pada *author* Hanim Maria Astuti, untuk mengetahui apakah keduanya memang saling berhubungan berdasarkan distribusi topiknya atau tidak. Distribusi topik dari *author* Hanim Maria Astuti sesuai dengan

Tabel 6.6.3.

Tabel 6.6.3 Distribusi topik *author* Hanim Maria Astuti

Hanim Maria Astuti	
Topik 40	0.95011
Topik 46	0.04943

Diketahui bahwa dokumen 3 dan *author* Hanim Maria Astuti memang berhubungan secara topik yang mendefinisikannya, yaitu pada Topik 40. Dari daftar topik yang dibentuk oleh model, diketahui kata-kata yang terdistribusi dalam Topik 40 adalah ‘kelola’, ‘ti’, ‘tata’, ‘tata_kelola’, ‘organisasi’, ‘bisnis’, ‘tingkat’, ‘level’, ‘nilai’, dan ‘matang’.

Setelah mengetahui kemiripan antara dokumen 3 dengan *author* Hanim Maria Astuti, dilakukan pula pengecekan distribusi topik *author* asli dari dokumen 3. Distribusi topik *author* dokumen 3 sesuai dengan Tabel 6.6.4.

Tabel 6.6.4 Distribusi topik *author* Eko Nugroho

Eko Nugroho	
Topik 11	0.67262

Topik 15	0.27179
Topik 41	0.05155

Berdasarkan Tabel 6.6.4 diketahui bahwa distribusi topik *author* Eko Nugroho pada model memang tidak mirip dengan distribusi topik dokumen 3.

Terdapat banyak faktor mempengaruhi hasil pengujian menjadi seperti yang didapatkan oleh eksperimen yang dilakukan oleh tugas akhir ini. Salah satunya adalah ketidakkonsistenan *author* dalam melakukan penelitian di bidangnya. Hal ini terjadi pada kasus dokumen pengujian 1 dengan judul “Analisa Kesuksesan Sistem Informasi Akademik (SIKAD) di Perguruan Tinggi dengan menggunakan D & M IS Success Model (Studi Kasus: ITS Surabaya)”. Penulis dokumen pengujian ini adalah Ardhini Warih Utami dan Febriliyan Samopa. Namun sesuai dengan Gambar 6.6.3, nama *author* dokumen 1 justru tidak muncul dalam hasil pengujian.

Model Stemming 50 Topik: Dokumen 1

	Author	Score	Size
88	Annis Paramita Dilla	0.603198	1
130	Bayu Kelana	0.599306	1
261	Febrianto	0.597287	1
362	Irna Yuniar	0.597284	1
701	Wan Maritul Kifti	0.597046	1
280	Ger Mustofa	0.597024	1
506	Nurmaini Dalimunthe	0.596988	1
167	Deni Arifianto	0.596848	1
600	Sejati Waluyo	0.596503	1
374	Joko Lianto Buliali	0.596332	2

Gambar 6.6.3 Hasil pengujian dokumen 1

Dapat dilihat pada Gambar 6.6.3, nama teratas yang muncul adalah *author* Annis Paramita Dilla. Untuk mengetahui hubungan keduanya, dilakukan pengecekan distribusi topik pada dokumen 1 dan *author* Febriliyan Samopa, serta *author* Annis Paramita Dilla sebagai *author* yang dianggap paling mirip dengan dokumen 1.

Tabel 6.6.5 Distribusi topik dokumen 1

Dokumen 1	
Topik 29	0.00259
Topik 33	0.00170
Topik 14	0.00136
Topik 10	0.00118
Topik 25	0.00103

Pada Tabel 6.6.5 diketahui bahwa topik-topik pada dokumen 1 yaitu Topik 29, Topik 33, Topik 14, Topik 10, dan Topik 25. Kemudian topik-topik yang menggambarkan *author* asli dokumen 1 sesuai dengan Tabel 6.6.6, yaitu Topik 1, Topik 2, dan Topik 46.

Tabel 6.6.6 Distribusi topik *author* Febriliyan Samopa

Febriliyan Samopa	
Topik 1	0.42372
Topik 2	0.55402
Topik 46	0.02186

Berdasarkan perbandingan kedua distribusi topik ini diketahui bahwa dokumen 1 dengan distribusi topik pada *author* Febriliyan Samopa yang telah dibentuk oleh model tidak saling berhubungan.

Tabel 6.6.7 Distribusi Topik *author* Annis Paramita Dilla

Annis Paramita Dilla	
Topik 14	0.07962
Topik 24	0.08469
Topik 29	0.42727
Topik 33	0.29202
Topik 46	0.10604

Kemudian dilakukan pengecekan distribusi topik *author* Annis Paramita Dilla, sebagai *author* yang dianggap paling mirip oleh model. Hasil distribusi topiknya dapat dilihat pada Tabel 6.6.7, dan dari tabel ini diketahui bahwa dokumen 1 dan *author* Annis Paramita Dilla memiliki kesamaan topik yaitu Topik 14, Topik 29, dan Topik 33.

Untuk memastikan hasil ini, dilakukan perbandingan dokumen 1 dengan dokumen yang dibuat oleh *author* Annis Paramita Dilla. Pada Tabel 6.6.8 dapat dilihat bahwa dokumen 1 membahas tentang permodelan dan pengolahan data menggunakan *Structural Equation Modeling*.

Tabel 6.6.8 Dokumen 1

Analisa Kesuksesan Sistem Informasi Akademik (Siakad) di Perguruan Tinggi dengan Menggunakan D & M Is Success Model (Studi Kasus: ITS Surabaya)

Ardhini Warih Utami, Febriliyan Samopa

Abstrak – Implementasi sistem informasi akademik dalam lingkungan perguruan tinggi digunakan untuk pengolahan data akademik. Tujuan penelitian ini adalah mengukur dan menguji variabel-variabel serta indikator yang mempengaruhi kesuksesan sistem informasi akademik di perguruan tinggi dengan menggunakan perguruan tinggi ITS Surabaya sebagai studi kasus. Model penelitian yang dipergunakan mengadopsi model penelitian Delone & Mclean IS Success. satu set kuesioner yang disusun dalam penelitian disebarkan kepada 115 responden yang memiliki kualifikasi sebagai pengguna sistem informasi akademik ITS (mahasiswa, dosen, staff karyawan). Uji validitas, reliabilitas dan outlier terhadap data penelitian dilakukan untuk memperoleh data yang valid dan reliabel. Selanjutnya data diolah dengan menggunakan Structural Equation Modelling. Pada penelitian ini terdapat 2 model penelitian, variabel yang dipergunakan dalam penelitian ini antara lain kualitas sistem, kualitas informasi, kualitas layanan, pemakaian, kepuasan pengguna dan manfaat-manfaat bersih. Hasil penelitian menunjukkan bahwa pada model 1 terdapat korelasi yang terjadi antar variabel yaitu kualitas sistem, kualitas informasi dan kualitas layanan memiliki pengaruh positif yang signifikan terhadap kepuasan pengguna sistem informasi akademik. Kualitas layanan dan kepuasan pengguna juga memiliki pengaruh positif yang signifikan terhadap pemakaian sistem informasi akademik, pemakaian sistem informasi akademik dan kepuasan pengguna sistem informasi akademik memiliki pengaruh positif yang signifikan terhadap manfaat-manfaat bersih sistem informasi akademik. Pada model 2 juga terdapat hubungan antarvariabel yaitu: Kualitas sistem dan kualitas layanan memiliki pengaruh positif yang signifikan terhadap pemakaian sistem informasi akademik. Kualitas sistem, informasi, layanan dan pemakaian sistem informasi akademik juga memiliki pengaruh positif yang signifikan terhadap kepuasan pengguna sistem informasi akademik. Pemakaian sistem informasi akademik dan kepuasan pengguna sistem informasi akademik memiliki pengaruh positif dan signifikan terhadap terhadap manfaat-manfaat bersih sistem informasi akademik. Berdasarkan hasil analisa, kesuksesan sistem informasi akademik di ITS memiliki prosentase sebesar 62 % yang artinya kesuksesan sistem informasi akademik ITS berada pada tingkatan "sukses" dengan didukung oleh faktor-faktor kesuksesan yang dihasilkan dalam penelitian.

Lalu pada Tabel 6.6.9, diketahui bahwa dokumen yang ditulis oleh *author* Annis Paramita Dilla membahas penggunaan *Technology Acceptance Model* juga dengan teknik *Structural Equation Modeling*.

Tabel 6.6.9 Dokumen *author* Annis Paramita Dilla

Analisis Keberhasilan Implementasi Rail Ticket System Menggunakan Pendekatan *Technology Acceptance Model*
Annis Paramita Dilla, Bambang Setiawan

Abstrak – Isu terhambat dalam pengelolaan SI/TI adalah sulitnya mengidentifikasi faktor yang mempengaruhi pengguna dalam menerima dan memanfaatkan suatu sistem. Rail Ticket System (RTS) merupakan sistem pemesanan tiket online yang baru diterapkan oleh PT. KAI. Tantangan PT. KAI saat ini adalah bagaimana mengembangkan RTS agar dapat diterima dan dimanfaatkan oleh masyarakat dalam jangka waktu yang panjang. Maka perlu dilakukan analisis penerimaan pengguna terhadap RTS menggunakan *Technology Acceptance Model* (TAM) dengan teknik *Structural Equation Modeling* (SEM) sebagai metode pengujian model. Hasil yang didapatkan dari penelitian ini adalah masyarakat Kota Surabaya telah dapat menerima RTS sebagai alternatif pemesanan tiket kereta api, dengan faktor yang mempengaruhi penerimaan pengguna adalah kemanfaatan sistem dan kemudahan pengguna.

Hal ini sesuai dengan kemiripan dokumen uji 1 dan *author* Annis Paramita Dilla berdasarkan distribusi topiknya dan menjelaskan mengapa perhitungan similaritas vektor probabilitas menghasilkan Annis Paramita Dilla sebagai *author* yang paling mirip.

Kasus di atas menjelaskan bahwa prediksi *author* menggunakan metode ini membutuhkan data yang konsisten dan menggambarkan masing-masing *author* dengan baik dan distribusi jumlah dokumen yang seimbang. Setiap *author* yang menulis lebih dari 1 dokumen, dibutuhkan dokumen pelatihan yang memang benar-benar menggambarkan bidang penelitian *author* tersebut secara spesifik, sehingga tidak menghasilkan topik-topik penelitian *author* yang terlalu luas dan melebar. Dengan begini, hasil prediksinya akan lebih konsisten.

6.6.3 Validasi *Keyword* dengan Distribusi kata dalam Topik

Dalam sebuah penelitian, *keyword* merupakan salah satu hal yang dianggap penting oleh penulis dan dapat dikatakan menggambarkan pokok bahasan dari penelitian tersebut secara umum. Oleh karena itu, berdasarkan hasil vektor probabilitas dokumen uji yang didapatkan dari distribusi kata-katanya, dilakukan validasi antara distribusi kata-kata topik yang muncul sebagai vektor probabilitas dokumen dengan *keyword* yang dimiliki oleh setiap dokumen untuk mengetahui kemampuan prediksi model terhadap *keyword* dari setiap dokumen uji. Persentase kemampuan prediksi model sesuai dengan Tabel 6.6.10.

Tabel 6.6.10 Recall *keyword* dokumen pengujian

Dokumen Pengujian	Persentase Prediksi Model	Dokumen Pengujian	Persentase Prediksi Model
1	67%	11	0%
2	33%	12	67%
3	67%	13	67%
4	50%	14	0%
5	40%	15	25%
6	33%	16	67%
7	33%	17	33%
8	50%	18	0%
9	67%	19	25%
10	17%	20	20%
Rata-rata persentase kemampuan prediksi model: 38.05%			

Pada Tabel 6.6.10 diketahui bahwa terdapat enam dokumen dengan persentase tertinggi 67%, yaitu dokumen 1, 3, 9, 12, 13, dan 16. Dari 3 *keyword* yang dimiliki oleh masing-masing dokumen ini, model dapat memprediksi dua diantaranya dengan benar. Sedangkan terdapat tiga dokumen yang tidak dapat diprediksi oleh model sama sekali dengan persentase 0%, yaitu dokumen 11, 14, dan 18. Secara keseluruhan, dari 20 dokumen

yang digunakan untuk melakukan pengujian model, diketahui bahwa rata-rata persentase kemampuan prediksi model adalah 38.05%.

Hal ini terjadi karena perbedaan distribusi kata-kata yang dimiliki setiap dokumen. Misalnya dokumen dengan *keyword* algoritma *Naïve Bayes*. Kata 'algoritma' merupakan kata yang memiliki frekuensi cukup tinggi di beberapa dokumen dengan topik serupa, sehingga model mengambil kata 'algoritma' menjadi kata yang menyusun distribusi salah satu topik. Sedangkan kata '*Naïve Bayes*' hanya muncul pada beberapa dokumen pelatihan saja, tidak sebanyak kata 'algoritma', sehingga model tidak mengenali kata '*Naïve Bayes*' sebagai 10 kata teratas yang dapat menyusun distribusi salah satu topik. Sehingga pada distribusi kata dalam topik, kata 'algoritma' muncul, namun kata '*Naïve Bayes*' tidak. Berdasarkan hal ini, dibutuhkan perlakuan khusus untuk menyaring *keyword* agar dapat mempengaruhi distribusi kata-kata yang menyusun topik yang dihasilkan oleh model.

Halaman ini sengaja dikosongkan

BAB VII

KESIMPULAN DAN SARAN

Pada bab ini dibahas mengenai kesimpulan dari semua proses yang telah dilakukan dan saran yang dapat diberikan untuk pengembangan yang lebih baik.

7.1 Kesimpulan

Kesimpulan yang didapatkan dari proses pengerjaan tugas akhir yang telah dilakukan antara lain:

1. Tahapan pra-proses data memiliki pengaruh yang cukup signifikan pada pembentukan model, dibuktikan dengan perbedaan hasil penentuan jumlah topik yang terbentuk dengan jenis pra-proses yang juga berbeda-beda, yaitu data tanpa *stemming*, dengan *stemming*, dan *noun POS Tagging*.
2. Data dengan *stemming* menghasilkan model yang lebih baik dalam memrediksi *author* yang menulis dokumen pengujian yang digunakan. Hal ini dibuktikan dengan persentase pada skenario *stemming* adalah 25%, dimana persentase ini yang paling baik dibandingkan dengan skenario lainnya.
3. Jumlah topik yang dapat memberikan hasil prediksi untuk keperluan rekomendasi *reviewer* dengan konsisten adalah model dengan jumlah topik 50 pada skenario *stemming*, dengan nilai *perplexity* 128.71 dan nilai *topic coherence* - 1.528.
4. Nilai *minimal author size* yang optimal dalam pengujian pada data yang digunakan di tugas akhir ini adalah 1, untuk menghasilkan prediksi yang lebih sesuai, berdasarkan rata-rata jumlah dokumen yang ditulis oleh *author* adalah 1.288 dan standar deviasi 0.999.

7.2 Saran dan Penelitian Selanjutnya

Dari pengerjaan tugas akhir ini, adapun beberapa saran untuk pengembangan penelitian ke depan.

1. Dibutuhkan data dengan *author* yang melakukan penelitian dengan konsisten untuk dapat menghasilkan pembelajaran model yang lebih baik sehingga dapat menghasilkan kemampuan prediksi model yang lebih baik.
2. Data pelatihan dipilih berdasarkan topik spesifik *author* untuk menghasilkan topik-topik dengan *author* yang lebih jelas.
3. Memisahkan *keyword* dari masing-masing penelitian untuk tidak diproses sama seperti teks keseluruhan melalui tahap *stemming* dan tokenisasi, sehingga tetap menjadi satu frasa.
4. Melakukan pelatihan data berdasarkan penelitian *author* dalam bahasa inggris yang ditranslasi.
5. Untuk diteliti lebih lanjut, dapat menggunakan data yang lebih terdistribusi secara merata dari sisi jumlah dokumen yang ditulis oleh *author*.
6. Melakukan pendefinisian kelompok topik berdasarkan model yang telah dihasilkan dengan pengamatan dan analisis yang lebih mendalam mengenai topik-topik penelitian yang telah ada dalam repositori *OAJIS* dengan membuat hirarki topik berdasarkan banyaknya jumlah topik yang ditemukan.

DAFTAR PUSTAKA

- [1] R. E. T. Al, P. Smyth, M. Steyvers, M. Rosen-zvi, U. C. Irvine, and U. C. Irvine, "The Author-Topic Model for Authors and Documents," 2004.
- [2] M. Steyvers, P. Smyth, and M. Rosen-zvi, "Probabilistic Author-Topic Models for Information Discovery," no. 1990, pp. 306–315, 1999.
- [3] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," vol. 3, pp. 993–1022, 2003.
- [4] Z. Xu, R. Lu, L. Xiang, and Q. Yang, "Discovering User Interest on Twitter with a Modified Author-Topic Model," 2011.
- [5] E. Linstead, P. Rigor, S. Bajracharya, C. Lopes, and P. Baldi, "Mining Eclipse Developer Contributions via Author-Topic Models," pp. 5–8, 2007.
- [6] F. Rashel, A. Luthfi, A. Dinakaramani, and R. Manurung, "Building an Indonesian Rule-Based Part-of-Speech Tagger."
- [7] G. G. Chowdhury, "Natural language Processing," pp. 51–89.
- [8] E. D. Liddy, "Natural Language Processing," 2001.
- [9] N. indurkha and F. J. . Damerau, *NATURAL LANGUAGE PROCESSING* .
- [10] A. Chopra, A. Prashar, and C. Sain, "Natural Language Processing," vol. 1, no. 4, pp. 131–134, 2013.
- [11] D. M. Blei, "Probabilistic Topic Models," pp. 77–84.
- [12] M. Adriani, "Probabilistic Part Of Speech Tagging for Bahasa Indonesia."
- [13] J. Chang *et al.*, "Reading Tea Leaves : How Humans Interpret Topic Models," 2009.
- [14] D. O. Callaghan, D. Greene, J. Carthy, and P. Cunningham, "Expert Systems with Applications An analysis of the coherence of descriptors in topic modeling," *Expert Syst. Appl.*, vol. 42, no. 13, pp. 5645–5657, 2015.

- [15] S. Zhu, “Information Retrieval using Hellinger Distance and Sqrt-cos Similarity,” no. Iccse, pp. 925–929, 2012.
- [16] M. Rosen-zvi and T. Griffiths, “Learning Author-Topic Models from Text Corpora,” vol. 28, no. 1, 2010.
- [17] O. Mortensen, “The Author-Topic Model Olavur,” 2017.
- [18] Gensim, “The Author-Topic Model: LDA with Metadata,” p. 223.
- [19] A. Purwarianti, A. Andhika, A. F. Wicaksono, I. Afif, and F. Ferdian, “InaNLP : Indonesia Natural Language Processing Toolkit Case study: Complaint Tweet Classification,” pp. 5–9, 2016.

BIODATA PENULIS



Penulis lahir di Yogyakarta pada tanggal 27 Mei 1996. Penulis merupakan anak pertama dari 2 bersaudara. Penulis telah menempuh beberapa pendidikan formal yaitu, SDN 16 Mataram, SMP Negeri 2 Mataram, dan SMA Negeri 1 Mataram.

Pada tahun 2013 pasca kelulusan SMA, penulis melanjutkan pendidikan dengan jalur SBMPTN (Tulis) di Jurusan Sistem Informasi FTIK – Institut Teknologi Sepuluh Nopember (ITS) Surabaya dan terdaftar sebagai mahasiswa dengan NRP 5213100110. Selama menjadi mahasiswa, penulis mengikuti berbagai kegiatan kemahasiswaan seperti beberapa kepanitiaan serta pernah menjabat sebagai Staf Departemen *Student Resource Development* BEM FTIf ITS dan pada tahun ketiga menjabat sebagai *Manager* Departemen *Student Resource Development* BEM FTIf ITS. Di bidang akademik, penulis aktif menjadi asisten dosen mata kuliah Interaksi Manusia dan Komputer. Selain itu, pada Bulan Oktober 2016 hingga Februari 2017 penulis berkesempatan menjadi penerima *JASSO Scholarship* untuk melakukan pertukaran pelajar pada program *Junior Year Program in English (JYPE)* di *Tohoku University*, Sendai, Jepang.

Pada tahun keempat, karena penulis memiliki ketertarikan di bidang *Natural Language Processing*, maka penulis mengambil bidang minat Akuisisi Data dan Diseminasi Informasi (ADDI). Penulis dapat dihubungi melalui *email* di kirunijah@gmail.com.

Halaman ini sengaja dikosongkan

LAMPIRAN A

Daftar hasil tugas akhir sesuai Tabel 1 dapat diakses pada <https://intip.in/repositoriADDI2018>

Tabel 1 Daftar repositori hasil tugas akhir

No.	Hasil Tugas Akhir	Keterangan
1.	Daftar <i>Stopwords</i>	Daftar <i>stopword</i> yang digunakan untuk melakukan tahapan pra-proses data pada tugas akhir ini
2.	Daftar Frasa Kata	Daftar frasa kata dari data pelatihan <i>OAJIS</i> yang digunakan untuk melakukan tahapan <i>POS tagging</i> pada tugas akhir ini.

A-1. Hasil Eksperimen Penentuan Jumlah Topik berdasarkan Nilai *Perplexity* – Tanpa *Stemming*

Percobaan	Jumlah Topik														
	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150
1	180.4	174	173.9	173.1	169.6	170.4	172.1	171.7	171.2	169.7	172.6	170.6	173.5	175	174.3
2	177.9	174.9	173.7	170.6	170.3	169.6	170.4	171	171.8	171	173.5	171	174.8	173.6	175.2
3	176.5	175.5	173	169.7	172	170.6	170.8	171.6	169.5	169.9	172.3	175.6	174.1	171.8	175.7
4	180.7	175.8	171.4	172.3	171.9	170.3	171.4	172.7	170.4	172.6	173.1	172.7	175.1	174.3	175.5
5	179.3	173.2	172.1	173.3	171.9	171.2	169.2	171.6	171.4	172.4	171.6	169.9	170.8	174.3	174.8
6	179.6	174.9	172.6	171.7	171.3	170.4	170.7	170.4	172.8	171.8	171.5	172.3	175.2	173.5	175.9
7	180.3	173.3	171.7	171.1	172.5	172.4	171.8	171	173.9	173	174.2	173	175	173.2	174.3
8	179.9	176.1	173.5	171.6	170.4	170.4	170.1	171.6	169.5	170.5	173.7	174.2	174.2	172.9	176.4
9	180.8	176.1	172.8	170.9	169.4	169.4	172.8	171.2	170.6	174.2	172.9	174.8	172.3	175.8	175.1
10	179.7	175.5	174	172.3	173.3	173.1	171.8	169.7	172.2	173.5	172.4	171.3	175.2	174.5	174.8
11	178.7	173.7	173.2	173.5	171.7	171.8	171.4	169.2	170.3	171.1	171.4	172.2	175	175.4	172.4
12	181.7	176.1	173.4	170.7	170.4	171.6	171.4	171.9	171.7	171.4	170.5	174.2	172.5	173.4	174.6
13	180.7	174.8	171.7	172.9	171.3	171.9	170.5	169.7	169.3	172.5	170.1	173.9	175.7	174.2	173.7
14	179.4	175.4	173.2	173.3	170	172.6	172.3	171.6	172.3	168.4	172.1	174.6	175.9	170.4	174.4
15	179.4	175.4	173.3	171.7	172.3	169.4	170.2	170.8	170.5	172.9	172.4	172.5	175.8	173.9	177.8
16	180.8	175	173.8	170.7	172.4	172.8	173	171.3	171.6	174.4	171.3	174	174.1	174.1	174.9
17	177.8	175.8	173.6	172.1	173.6	171.9	172.8	171.8	173.3	173.3	173.5	175.4	173.9	175.6	175.6
18	181.7	174	173.1	171.1	169.5	172.5	171.1	172.5	171.3	173.2	171.8	172.3	170.4	173.3	175.5
19	178.4	174.3	174.2	173.4	172	173.4	172.3	172	173.8	172.2	171.2	172.4	173.8	174.8	175.2
20	180.1	174.3	173.6	172.3	169.6	171.3	171.4	172.3	171.1	171.1	174	172.5	173.3	176.6	174.1
21	179.5	175.5	173.8	172.2	170.8	173	171.5	172	171.7	170.5	175.2	174.4	174	176.2	174.7
22	179.7	175.5	172	172.3	169.7	171.6	172.4	172.6	171.3	171.1	174.7	173.7	173.6	176.5	171.8
23	181.4	176.2	172.1	171	171.9	170.8	171.5	170.9	170.9	169.7	171.6	174.6	172.1	173.3	176.6
24	181.4	173.6	174.2	169.2	170.6	173.5	170.5	170.4	171.5	173	173	174.5	175.8	176.2	178.1

25	179.9	172.5	173.9	170.3	171.9	171.4	171.6	171.2	171.8	172.3	171.8	172.2	172.6	175.1	176
26	176.5	174.4	173.2	173	171	169.1	169.9	170.2	170.9	170	168.4	175.2	177	174.2	174.5
27	177.9	174.3	171.9	171.7	171.9	171.7	172.1	170	172.9	170.5	170.1	173.4	173	174.7	174.7
28	178.8	174.4	172.6	172.7	169.8	172.7	169.3	173.8	171.8	171.7	173.5	173.1	174.4	175.4	172.3
29	181.5	175.9	172.9	170.6	170.5	170.5	171.5	170.7	170.9	170.9	172.3	171.6	174.7	172.2	176.4
30	177	172.9	171.1	173.1	171.8	172.2	168.5	172.8	169.7	171.1	171.1	172.5	171.5	172.7	175.6

A-2. Hasil Eksperimen Penentuan Jumlah Topik berdasarkan Nilai *Perplexity* – *Stemming*

Percobaan	Jumlah Topik														
	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150
1	137.6	132.9	131.3	128.4	127.5	128.4	128.9	126	129	127.1	128.4	128.9	127.6	127.7	129.6
2	138.7	133.6	130.1	129.8	127.9	128.4	128.9	128.3	126.8	127.3	127.4	128.3	128.9	128.8	128.2
3	140.3	131.8	132.3	127.6	128.2	128.5	126.4	128	127.6	127.7	127	130.2	128	128.9	130.1
4	138	133.1	132.5	128.3	129.1	129.1	127.7	127.4	127.7	127.1	128.7	127.7	126.6	128.1	127.2
5	136	132.8	131.6	128.5	129.1	128.1	128	127.8	127.8	126.9	126.5	128.2	127.1	127.9	129.2
6	138.1	131.7	131.4	129.2	128.6	129.1	127.4	127.1	127.3	129.6	126.8	128.4	127.8	129.6	127.7
7	138.6	132.6	130.3	130.6	126.6	127.5	128.3	128.7	128.5	129.6	128.2	128.6	128.1	128.4	129.3
8	138.5	132.5	130.5	130.5	128.7	129.6	126.6	126.2	125.4	128	129.6	127.9	128.4	128.6	127.5
9	138	133.6	130.5	129.7	128.4	127.2	126.9	128.2	127.8	126.4	127.7	128.1	128.4	129.2	127.6
10	139.1	133.9	130.1	129.8	129.9	127.8	129	129	127.9	126.3	127.9	128.1	128.3	126.1	128.2
11	138.8	131.9	131.3	128.9	128.8	127.5	127.5	127.9	127.2	126.2	127.2	129.1	128	128.1	128.4
12	139.4	133.1	131	128.6	128.5	127.4	127.1	128	127.3	127.3	128.6	127.5	128.3	128.9	130.6
13	136.5	134	131.4	129	128.2	127.7	128.2	127.6	127.3	129.3	127.5	129.1	128.5	128.7	127.4
14	137.8	132.8	130.7	129.1	129.8	127.2	127.3	126.7	127.6	126.7	127.7	127.9	128.8	130	128.9
15	137.4	133.7	132	130.4	129.9	127.8	129.2	127.6	127.6	126.7	126.5	128.8	128.4	128.8	127.4
16	138.4	133.8	130.7	129.3	128.7	128.8	127.2	129.2	127.8	127.2	127.8	127.3	128	127.5	127.5
17	135.9	131.8	131.7	130.7	129.7	128.5	127.4	128.3	127.6	126.1	127.1	127.5	126.2	128.2	129
18	138.2	133.5	130.4	129.2	130.4	125.4	128.3	126.8	128.1	127.3	126.3	127.3	128.6	127.6	127.1
19	137.8	132.5	129.8	128.3	127.4	127.4	128.7	128.5	128.9	126.9	128.5	128.5	128.9	128.8	129.6
20	136.8	133.6	128.8	128.3	128.4	127.7	127.3	126.7	127.8	127.6	126.4	129.7	127.7	125.5	130.1
21	138.2	133.7	131.4	129.1	130.1	128.5	127.5	125.6	129	128.2	127.4	127.8	127.2	128.1	128.9
22	136.5	132.4	130.1	128.5	127.8	127.6	128	128.3	127.1	128	128	129.3	128.1	126.9	128.1
23	140.5	135.4	131.1	129.8	127.5	127.9	127.1	126.3	125.6	128.2	127.1	128	128.1	127.9	130.4
24	136.1	134.5	131.9	130.2	129.6	127.4	125.9	127.5	127.7	128.1	127.3	127.1	128.3	129	127.3

25	137.5	132.3	130.1	130.1	128.3	127.7	127.6	127.5	127.4	126.9	127.5	128.4	128.2	126.8	129.5
26	137.2	131.5	131.8	128.2	128	129.4	127.2	128	127.2	128.9	127.8	127	127	126.9	128.7
27	139.1	132.6	131.9	128.1	128.8	129.1	128.3	129	128.2	125.1	128.6	128.5	128.2	129	127.1
28	137.3	132.5	131.7	129.9	128.9	129.5	126.4	128.1	128.2	127.3	126.5	127.4	129.4	129.9	128.4
29	136.9	133.7	132	129	129.3	127.6	127.9	127.3	127.5	127.6	126.7	127.7	126.7	129.1	127.6
30	136.8	132.3	129.4	129.8	129.3	128.6	127.7	127.6	128	128.2	126.7	128	129.6	126.7	127.8

A-3. Hasil Eksperimen Penentuan Jumlah Topik berdasarkan Nilai *Perplexity* – *Noun POS Tagged*

Percobaan	Jumlah Topik														
	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150
1	97.7	93.4	91.1	89	88.1	86.9	87	85.4	86.5	85.9	85.7	84.9	86.1	85.9	85.4
2	96.4	93.1	91.6	89.9	88.3	87.4	87.4	86	86.4	85.9	86.3	86	84.6	86	84.6
3	98.3	94	90.8	88.7	88.7	88.7	85.8	85.4	85.1	84.9	85.2	85.3	85.4	86.3	83.7
4	100.4	92.1	90.1	90	89.4	87.8	87.5	85.4	86.1	86.6	85.7	84.5	86	84.8	84.2
5	95.7	93.4	88.8	89.3	88	87.5	85.4	85.7	86	85.5	86	84.3	86.2	85.5	84.3
6	96.8	93.2	90.9	90.2	88	87.6	87	85.9	86.1	86.6	84.5	84.9	85.5	85.4	85.3
7	99.1	94.1	91.1	90.3	88.4	86.1	86.1	86.8	86	85.8	85.3	84.6	85.2	85.5	84.6
8	98.4	94.6	91.4	90.5	88.3	87.5	85.9	85.1	86.1	85.9	85.1	84.6	84.6	85.9	84.9
9	98.8	94.1	91.2	89.1	87	87.3	87.4	86.9	85.4	85.3	84.2	86.4	85	85.1	84.7
10	97.3	94.6	90.7	89.3	90.3	87.2	86.8	86.3	86.8	86.1	84.9	86.2	85	85.2	85.2
11	99.1	92.9	90	89.4	87.9	86.5	87.1	85.5	84.7	84.9	84.9	84.9	84.4	84.6	85.3
12	98.2	94	91.5	89.9	88	88.1	86.5	86.6	86.6	87.1	85.3	85.9	87	84.7	85.4
13	99.5	92.4	90.9	88.7	88.7	87.5	86.3	86.7	85.8	86.1	85.8	85.3	85.5	84.2	84.8
14	99.1	94.5	90.8	90	88.4	88.1	86.2	87.8	86.1	86.4	86	86.1	85.1	85.2	86.5
15	97.7	93.2	90.5	89.7	88.4	87.7	86.8	87.1	86.4	85.6	85.5	85.1	86.2	84.3	84.7
16	99.2	93.6	92.2	89.4	87.9	87.5	87.8	85.4	85.5	86.1	86.6	86.4	86.1	84.3	86.4
17	96.1	92.7	92.8	89.6	86.5	88.1	86.4	86.7	84.5	85.2	86	85.7	84.3	84.4	84.9
18	98.5	93	90.9	89.8	88.4	85.8	87.7	86.4	86.2	86.5	84.7	86.1	85.3	85.6	85.8
19	96.5	93.5	89.8	91.3	88.5	87.6	87.8	87.1	85.8	84.6	84.8	85.7	85.8	86	84.6
20	97.1	94	90.7	90.1	87.6	87.9	86.2	85.9	86.1	86	85.6	85.8	85.8	86.5	85.5
21	97.3	93.1	91.3	88.4	88.9	87.4	85.7	88	84.9	86.5	85.7	86.2	85.6	85	85.7
22	97.5	92.9	90.7	89.4	88.5	86.8	86.6	85.7	86.6	86.6	86	84.6	84.6	85.4	85.2
23	99.5	93.1	91.7	90.6	88.9	87.6	86.8	86.3	85.2	85.8	86.1	86.5	84.6	85.7	85.4
24	97.3	93	91.6	88.7	86.6	87.2	87.2	86.9	85.6	85	86.7	84.7	84.3	85.6	85

25	98.7	94.2	91.3	89.3	89	89	86.9	86.1	85.6	86.2	85.5	85.8	85.7	84.4	85.3
26	97.1	94	89.7	89.2	88	88	86.4	84.8	86	84.7	85.3	86	84.2	84.8	85.2
27	98.7	93.8	90.9	89.1	89.3	86.9	85.9	86.1	85.8	86.3	85.6	84.8	84.8	84.8	86.5
28	98.5	93.9	89.7	88.9	88.7	87.3	87	86.1	86.2	86.7	85	85.1	85.8	84.7	86.2
29	98.2	92.3	90.2	89.1	89.6	87.4	86.6	86.8	86.2	85.6	85.7	85.4	85.9	85.1	84.8
30	97.1	93.7	91.8	89.2	88.8	87.9	86.8	85.9	86.3	86.7	85.5	84.7	85.8	84.9	84.9

Halaman ini sengaja dikosongkan

LAMPIRAN B

B-1

B-1. Tanpa Stemming: Daftar 50 Topik

0	1	2	3	4	5	6
0.026*diagram	0.028*risiko	0.018*uji	0.020*menu	0.039*aplikasi	0.011*pengguna	0.031*aplikasi
0.019*form	0.015*ti	0.017*pesanan	0.019*halaman	0.020*mobile	0.010*nilai	0.016*pengguna
0.019*barang	0.014*teknologi	0.016*metode	0.018*sim	0.011*program	0.009*user	0.015*kebutuhan
0.014*menu	0.013*kendaraan	0.015*file	0.017*card	0.010*pengguna	0.008*aktivitas	0.009*game
0.011*pelanggan	0.012*keamanan	0.014*kendaraan	0.013*file	0.010*tampilan	0.008*analisis	0.009*perangkat
0.010*penjualan	0.011*model	0.010*coba	0.012*pengujian	0.009*web	0.007*metode	0.008*user
0.009*makanan	0.011*pembelajaran	0.010*aplikasi	0.011*pelanggan	0.009*wisata	0.007*system	0.007*fungsi
0.009*case	0.011*kegiatan	0.009*sensor	0.010*usaha	0.008*berbasis	0.006*memiliki	0.007*permainan
0.009*sistem_informasi	0.009*aplikasi	0.009*code	0.009*bukti	0.007*objek	0.006*teknologi	0.007*pengujian
0.008*user	0.009*mahasiswa	0.008*uji_coba	0.008*digital	0.007*kegiatan	0.006*perangkat	0.006*virtual
7	8	9	10	11	12	13
0.022*laporan	0.068*pelanggan	0.015*penjualan	0.034*model	0.015*jadwal	0.026*siswa	0.023*erp
0.012*kriteria	0.029*customer	0.015*input	0.030*nilai	0.015*akses	0.015*aplikasi	0.019*perusahaan
0.011*kerja	0.027*bit	0.013*web	0.025*tanaman	0.015*tabel	0.013*dokumen	0.015*kriteria
0.011*keputusan	0.025*iterasi	0.013*case	0.016*obat	0.014*support	0.011*nilai	0.011*aplikasi
0.010*sistem_informasi	0.024*matriks	0.011*perusahaan	0.009*dimana	0.012*nilai	0.010*tabel	0.011*biaya
0.010*metode	0.022*peta	0.009*produk	0.009*parameter	0.011*dosen	0.008*memiliki	0.010*bisnis
0.010*tabel	0.020*ukuran	0.009*form	0.009*panas	0.011*rumah	0.008*sekolah	0.010*alternatif
0.010*nilai	0.018*decision	0.008*item	0.009*indeks	0.010*pattern	0.007*tahap	0.010*sistem_informasi
0.009*tenaga	0.017*check	0.008*anggota	0.009*bumi	0.010*sakit	0.007*guru	0.009*proyek
0.008*bobot	0.015*maksimum	0.007*jari	0.008*metode	0.009*item	0.007*indonesia	0.009*enterprise

14	15	16	17	18	19	20
0.022*sistem_informasi	0.033*pengguna	0.028*ti	0.022*model	0.035*obat	0.055*publik	0.034*fuzzy
0.020*dosen	0.023*variabel	0.027*nilai	0.011*nilai	0.024*resiko	0.049*aplikasi	0.014*metode
0.018*kinerja	0.023*nilai	0.022*bisnis	0.009*tabel	0.016*level	0.048*layanan_publik	0.010*jenis
0.012*pengolahan	0.019*faktor	0.020*keselarasan	0.007*sistem_informasi	0.015*mahasiswa	0.034*kota	0.010*keamanan
0.011*penilaian	0.018*penggunaan	0.019*faktor	0.007*pelanggan	0.015*elemen	0.028*layanan	0.010*clustering
0.010*evaluasi	0.012*model	0.017*strategi	0.007*rumah	0.012*bebas	0.024*pemerintah	0.008*kerja
0.009*satelit	0.011*indikator	0.014*rekomendasi	0.006*skenario	0.012*metode	0.021*pembangunan	0.008*pola
0.009*perangkat	0.009*analisis	0.012*level	0.006*kondisi	0.011*penilaian	0.021*pengembangan	0.008*suara
0.009*level	0.009*hipotesis	0.011*universitas	0.006*listrik	0.011*klasifikasi	0.019*government	0.008*simulasi
0.008*pengguna	0.009*layanan	0.011*xyz	0.006*user	0.011*manajemen	0.017*surabaya	0.008*himpunan
21	22	23	24	25	26	27
0.025*organisasi	0.034*pengetahuan	0.013*mahasiswa	0.060*gejala	0.031*fungsi	0.054*bahasa	0.033*model
0.020*ti	0.023*management	0.011*sistem_informasi	0.043*penyakit	0.031*metode	0.031*bali	0.025*skenario
0.020*kelola	0.022*service	0.009*kegiatan	0.039*nilai	0.024*fuzzy	0.027*aplikasi	0.022*produksi
0.019*tata	0.017*knowledge	0.008*kriteria	0.031*operasi	0.022*variabel	0.025*indonesia	0.022*website
0.017*tata_kelola	0.014*notasi	0.007*komputer	0.026*tabel	0.017*harga	0.022*anak	0.022*lahan
0.011*jaringan	0.013*fitur	0.007*keputusan	0.023*pakar	0.017*nilai	0.017*rakyat	0.022*jurusan
0.010*bisnis	0.012*problem	0.006*produk	0.022*hilang	0.016*tabel	0.017*bahasa_indonesia	0.018*konten
0.010*sistem_informasi	0.012*perusahaan	0.006*pemasaran	0.022*hash	0.016*coba	0.015*lokal	0.017*simulasi
0.007*model	0.012*karyawan	0.006*perguruan	0.022*kombinasi	0.015*analisis	0.014*koleksi	0.015*motor
0.007*strategi	0.011*resource	0.006*system	0.021*jenis	0.015*produk	0.012*piranti	0.014*kapasitas

28	29	30	31	32	33	34
0.033*uji	0.026*nilai	0.030*ti	0.034*algoritma	0.063*website	0.036*matriks	0.033*layanan
0.026*produksi	0.020*tabel	0.025*perusahaan	0.019*nilai	0.018*online	0.022*android	0.013*kebutuhan
0.026*mesin	0.019*metode	0.022*keputusan	0.017*model	0.017*pengunjung	0.016*algoritma	0.011*analisis
0.024*aplikasi	0.015*aktivitas	0.013*kinerja	0.014*bisnis	0.016*kategori	0.016*aplikasi	0.010*bisnis
0.015*peralatan	0.014*individu	0.012*penilaian	0.014*proses_ bisnis	0.016*produk	0.013*password	0.009*ti
0.015*pemeliharaan	0.013*peramalan	0.011*bank	0.013*citra	0.015*variabel	0.013*box	0.009*service
0.013*otentikasi	0.013*biaya	0.009*manajemen	0.009*pakar	0.014*halaman	0.011*login	0.009*kondisi
0.013*mysql	0.013*case	0.008*game	0.009*pelanggan	0.014*web	0.011*bit	0.008*teknologi
0.013*user	0.013*uji	0.007*item	0.008*log	0.013*node	0.010*input	0.007*tabel
0.012*perangkat	0.013*generasi	0.007*memiliki	0.008*aktivitas	0.013*penggunaan	0.010*berbasis	0.007*sop
35	36	37	38	39	40	41
0.030*rumah	0.015*tata	0.013*pemilihan	0.046*nilai	0.025*web	0.046*internet	0.022*video
0.019*sensor	0.014*tata_kelola	0.011*metode	0.021*fuzzy	0.016*berbasis	0.029*model	0.021*kebutuhan
0.012*kesehatan	0.014*kelola	0.008*dokumen	0.017*algoritma	0.016*pembelajaran	0.027*mahasiswa	0.017*diagram
0.011*sakit	0.013*dashboard	0.008*produksi	0.016*rendah	0.015*kinerja	0.023*penggunaan	0.017*kesehatan
0.010*pengujian	0.013*bisnis	0.008*pendaftaran	0.015*fungsi	0.013*uji	0.021*variabel	0.015*uc
0.010*memiliki	0.013*ti	0.008*suara	0.013*tabel	0.012*minat	0.020*cluster	0.015*desain
0.009*rumah_sakit	0.011*organisasi	0.007*diagram	0.009*akurasi	0.012*organisasi	0.020*belajar	0.014*pengguna
0.009*pengguna	0.010*kebutuhan	0.007*aplikasi	0.009*metode	0.011*berkas	0.020*nilai	0.012*perangkat
0.009*motor	0.010*teknologi	0.007*information	0.008*klasifikasi	0.011*pegawai	0.018*akademik	0.010*aplikasi
0.008*perangkat	0.009*proses_ bisnis	0.006*matriks	0.008*pakar	0.010*modul	0.015*penggunaan_ internet	0.009*lunak

42	43	44	45
0.030*portofolio	0.031*learning	0.048*sistem_ informasi	0.026*media
0.018*wisata	0.021*tingkat	0.041*kesehatan	0.013*dimensi
0.018*surat	0.018*teknologi	0.029*pelayanan	0.011*sosial
0.017*laporan	0.012*masarakat	0.027*web	0.011*social
0.014*kriteria	0.012*indonesia	0.025*penjualan	0.009*konsumen
0.011*rekomendasi	0.010*internet	0.023*fasilitas	0.008*tik
0.011*tingkat	0.010*kematangan	0.016*rujukan	0.008*organisasi
0.010*berkas	0.010*sms	0.012*transaksi	0.007*perusahaan
0.010*pengujian	0.010*cobit	0.011*database	0.007*ukm
0.009*sesuai	0.010*kesiapan	0.011*pelayanan_ kesehatan	0.007*media_sosial
46	47	48	49
0.019*dosen	0.024*prosedur	0.035*sistem_ informasi	0.045*koperasi
0.016*metode	0.021*kualitas	0.031*kualitas	0.023*aplikasi
0.014*ciri	0.020*perangkat	0.019*responden	0.019*multimedia
0.013*nilai	0.019*aktivitas	0.019*api	0.019*kesehatan
0.013*citra	0.017*harga	0.018*penilaian	0.017*tingkat
0.011*jaringan	0.017*dokumen	0.017*setuju	0.014*pembelajaran
0.009*input	0.014*standar	0.017*pernyataan	0.012*keuangan
0.009*tingkat	0.014*manajemen	0.013*sia	0.012*cabang
0.008*tabel	0.013*insiden	0.013*dimensi	0.011*materi
0.008*pengujian	0.012*lunak	0.013*menampilkan	0.011*bentuk

B-2. Stemming: Daftar 100 Topik

0	1	2	3	4	5	6
0.019*model	0.072*nilai	0.045*portofolio	0.221*main	0.034*akreditasi	0.053*tik	0.043*mysql
0.019*usaha	0.024*kerja	0.037*bahasa	0.084*game	0.031*nilai	0.031*kerja	0.027*query
0.018*dokumen	0.023*kriteria	0.030*indonesia	0.052*virtual	0.027*rendah	0.030*fungsi	0.027*cluster
0.018*modul	0.023*metode	0.023*bahasa_ indonesia	0.022*aplikasi	0.019*bebas	0.027*aplikasi	0.025*media
0.015*erp	0.018*bobot	0.018*uji	0.021*interaksi	0.019*online	0.026*ict	0.020*ajar
0.015*karyawan	0.015*prioritas	0.016*sedia	0.021*objek	0.017*dokumen	0.024*kota	0.018*aplikasi
0.015*diagram	0.014*putus	0.015*investasi	0.018*area	0.017*survei	0.023*organisasi	0.015*siswa
0.014*otentikasi	0.013*hitung	0.014*teks	0.017*run	0.015*rs	0.022*pimpin	0.015*sql
0.013*user	0.013*butuh	0.014*koleksi	0.017*tanda	0.015*logo	0.018*government	0.014*simpan
0.013*open	0.011*dukung	0.013*bentuk	0.017*nyata	0.014*jenis	0.017*rencana	0.014*tahap
7	8	9	10	11	12	13
0.081*obat	0.028*harga	0.059*rumah	0.040*individu	0.077*dashboard	0.098*tumbuh	0.030*framework
0.037*tanam	0.026*fuzzy	0.044*share	0.030*pilih	0.036*sensor	0.063*obat	0.018*fungsi
0.026*usaha	0.024*beli	0.031*aktor	0.025*optimal	0.035*motor	0.054*objek	0.014*file
0.023*beli	0.023*nilai	0.024*tahap	0.023*mahasiswa	0.033*login	0.043*lapis	0.014*aplikasi
0.016*jual	0.022*langgan	0.018*lihat	0.023*time	0.023*password	0.043*properti	0.014*identifikasi
0.016*transaksi	0.021*variabel	0.017*uji	0.020*generasi	0.021*fakultas	0.036*model	0.013*layan
0.015*catat	0.021*tingkat	0.016*motor	0.019*algoritma	0.020*mutu	0.033*aplikasi	0.012*web
0.014*atur	0.020*uji	0.016*actor	0.017*populasi	0.020*alat	0.033*definisi	0.011*lihat
0.013*aplikasi	0.017*coba	0.015*tugas	0.016*genetika	0.019*program	0.028*domain	0.010*ajar
0.013*gateway	0.016*produk	0.015*sesuai	0.016*distribusi	0.016*pin	0.027*identifikasi	0.010*ancang

14	15	16	17	18	19	20
0.087*citra	0.033*olah	0.082*sehat	0.076*resiko	0.060*algoritma	0.061*government	0.021*hash
0.034*kenal	0.020*satelit	0.052*sistem_ informasi	0.038*mahasiswa	0.042*nilai	0.051*faktor	0.020*makan
0.024*ciri	0.019*level	0.046*fasilitas	0.037*nilai	0.015*uji	0.045*kota	0.019*nilai
0.019*vektor	0.015*pilih	0.042*layan	0.029*manajemen	0.012*ciri	0.041*perintah	0.016*input
0.018*titik	0.015*olah_data	0.028*sedia	0.028*learning	0.010*genetika	0.038*sukses	0.015*fungsi
0.016*deteksi	0.014*citra	0.028*integrasi	0.027*terap	0.010*cari	0.032*implementasi	0.015*menu
0.016*matriks	0.014*measure	0.028*rujuk	0.027*teknologi	0.009*algoritma_ genetika	0.026*daerah	0.015*kerangka
0.015*pola	0.013*kualitas	0.018*layan_sehat	0.024*control	0.008*fitness	0.021*masyarakat	0.013*fungsi_ hash
0.014*karakter	0.013*metode	0.017*sumber	0.023*kesiap	0.008*peta	0.018*tingkat	0.013*pakar
0.012*metode	0.013*perangkat	0.017*daya	0.020*diskusi	0.008*minimum	0.017*dukung	0.013*kerja
21	22	23	24	25	26	27
0.053*sakit	0.057*monitoring	0.036*nilai	0.039*aplikasi	0.025*sistem_ informasi	0.041*langgan	0.042*layan
0.050*rumah	0.042*video	0.027*sistem_ informasi	0.029*warna	0.022*knowledge	0.036*sensor	0.040*uji
0.045*rumah_sakit	0.030*pratama	0.022*tuju	0.028*fitur	0.020*km	0.020*media	0.031*dokumen
0.030*sehat	0.027*commerce	0.021*kualitas	0.023*tangan	0.014*jaring	0.019*konsumen	0.020*klasifikasi
0.028*layan	0.020*wawancara	0.021*wisata	0.017*pola	0.013*bisnis	0.018*layan	0.016*email
0.016*sistem_ informasi	0.018*lapor	0.020*lihat	0.015*uji	0.013*strategi	0.016*node	0.015*coba
0.014*manfaat	0.015*ambil	0.019*uji	0.015*tulis	0.013*aplikasi	0.016*tingkat	0.015*uji_coba
0.013*pasien	0.015*unggah	0.019*tampil	0.015*tanda	0.012*guru	0.014*sosial	0.015*server
0.012*fungsi	0.015*sd	0.018*nyata	0.014*variabel	0.012*ti	0.014*jarak	0.015*buku
0.012*sedia	0.015*simpan	0.015*api	0.014*tanda_tangan	0.011*produk	0.012*media_sosial	0.015*kendara

28	29	30	31	32	33	34
0.081*sim	0.031*enterprise	0.034*kualitas	0.059*cluster	0.029*aplikasi	0.029*uji	0.014*nilai
0.073*card	0.030*rancang	0.028*tugas	0.036*calon	0.027*akun	0.025*tampil	0.012*learning
0.057*wisata	0.027*tabel	0.016*rate	0.031*pusat	0.016*orang	0.022*ajar	0.011*aman
0.034*objek	0.023*bisnis	0.016*jamin	0.031*pegawai	0.014*nilai	0.020*nilai	0.011*aktivitas
0.033*digital	0.023*lembaga	0.016*tabel	0.030*kriteria	0.014*koperasi	0.013*item	0.011*butuh
0.030*bukti	0.019*sistem_ informasi	0.016*aplikasi	0.030*metode	0.013*nama	0.011*putus	0.009*model
0.020*faktor	0.016*uji	0.014*verifikasi	0.027*program	0.012*profil	0.010*orang	0.009*fuzzy
0.020*telepon	0.015*basis	0.013*analisis	0.026*mahasiswa	0.012*giat	0.009*fk	0.008*tuju
0.019*barang	0.015*butuh	0.013*layan	0.025*putus	0.012*atribut	0.009*mata	0.008*amal
0.017*file	0.015*suara	0.013*butuh	0.023*pilih	0.012*user	0.009*kelas	0.007*faktor
35	36	37	38	39	40	41
0.049*ti	0.019*aplikasi	0.042*video	0.166*website	0.024*uji	0.029*layan	0.049*aplikasi
0.030*produksi	0.016*file	0.032*budaya	0.084*jurus	0.018*perangkat	0.011*ti	0.029*layan
0.028*adopsi	0.015*metode	0.021*barang	0.074*konten	0.017*aplikasi	0.011*teknologi	0.023*publik
0.026*mesin	0.010*tahap	0.020*uji	0.033*kategori	0.014*perangkat_ lunak	0.011*kerja	0.023*layan_ publik
0.024*analisis	0.010*pesan	0.020*id	0.032*peringkat	0.014*lunak	0.011*service	0.021*butuh
0.024*erp	0.010*pakar	0.019*form	0.025*rekomendasi	0.014*butuh	0.010*aplikasi	0.013*bangun
0.018*file	0.009*orang	0.019*varchar	0.020*universitas	0.013*tampil	0.010*tingkat	0.013*kota
0.018*bisnis	0.009*nilai	0.017*culture	0.015*teknik	0.011*implementasi	0.010*kondisi	0.011*surabaya
0.017*alat	0.009*lihat	0.017*uc	0.015*utara	0.010*diagram	0.009*management	0.008*sedia
0.016*aplikasi	0.009*tampil	0.016*akun	0.015*meta	0.010*lihat	0.009*akses	0.008*server

42	43	44	45	46	47	48
0.110*perceived	0.046*ajar	0.030*giat	0.040*website	0.025*nilai	0.091*ti	0.027*layan
0.064*model	0.035*mahasiswa	0.022*anggar	0.027*resource	0.019*metode	0.037*cobit	0.019*model
0.061*pengaruh	0.019*nilai	0.020*kerja	0.024*period	0.019*model	0.034*soft	0.016*usaha
0.053*kota	0.017*kuliah	0.019*uang	0.018*halaman	0.015*variabel	0.033*organisasi	0.013*butuh
0.047*musim	0.014*mata	0.017*sia	0.017*main	0.013*jaring	0.030*hard	0.013*metode
0.044*variabel	0.012*studi	0.015*lapor	0.015*cost	0.013*aman	0.029*level	0.012*tingkat
0.044*indeks	0.010*mata_kuliah	0.015*skpd	0.014*cari	0.013*uji	0.018*nilai	0.010*bisnis
0.041*rumah_sakit	0.010*online	0.013*tingkat	0.014*number	0.012*daerah	0.018*kompetensi	0.009*pengaruh
0.037*valid	0.009*tingkat	0.012*sistem_informasi	0.014*type	0.011*user	0.017*analisis	0.009*nilai
0.034*rumah	0.008*model	0.011*media	0.014*commerce	0.009*ukur	0.017*teknologi	0.009*putus
49	50	51	52	53	54	55
0.046*usability	0.051*model	0.042*aman	0.042*pustaka	0.033*kerja	0.077*insiden	0.097*kendara
0.031*rumah	0.048*akses	0.042*serang	0.041*non	0.030*tenaga	0.054*case	0.091*code
0.026*evaluasi	0.044*insiden	0.033*password	0.028*kondisi	0.030*tenaga_kerja	0.037*prosedur	0.050*baca
0.023*input	0.027*sop	0.019*authentication	0.026*orang	0.025*form	0.034*manajemen	0.038*ms
0.023*user	0.024*formulir	0.019*code	0.022*akuisisi	0.020*sistem_informasi	0.024*tangan	0.033*detik
0.022*sehat	0.023*admin	0.018*nilai	0.022*buku	0.018*diagram	0.022*diagnosa	0.033*identifikasi
0.020*distance	0.022*website	0.018*server	0.017*temperatur	0.017*jawa	0.020*solusi	0.028*coba
0.020*nilai	0.022*senjang	0.017*kompetisi	0.016*pakar	0.016*barat	0.018*bangkit	0.026*uji_coba
0.016*hitung	0.018*ar	0.017*team	0.016*atur	0.016*daftar	0.018*cerdas	0.022*klien
0.015*salah	0.018*bina	0.016*attack	0.015*citra	0.016*jawa_barat	0.017*memory	0.020*nomor

56	57	58	59	60	61	62
0.068*jual	0.032*kerja	0.048*butuh	0.047*layan	0.052*gejala	0.032*bisnis	0.051*tumpang
0.034*beli	0.018*organisasi	0.030*perangkat	0.038*kapasitas	0.045*sakit	0.025*proses_bisnis	0.023*rute
0.034*harga	0.016*usaha	0.021*perangkat_lunak	0.038*bank	0.031*pakar	0.024*usaha	0.021*tingkat
0.032*aplikasi	0.016*uji	0.021*lunak	0.028*android	0.029*game	0.021*level	0.021*jalan
0.027*mobile	0.015*pengaruh	0.014*simulasi	0.023*leadership	0.027*main	0.019*domain	0.020*lihat
0.026*learning	0.014*faktor	0.013*nilai	0.021*kriptografi	0.026*operasi	0.018*matang	0.019*block
0.026*produk	0.012*ti	0.013*power	0.019*perangkat	0.023*item	0.018*tingkat	0.018*skema
0.021*customer	0.009*pimpin	0.013*satelit	0.018*dukung	0.022*jenis	0.017*peta	0.018*ac
0.019*operator	0.009*lihat	0.012*anggota	0.017*daya	0.018*sistem_pakar	0.017*tingkat_matang	0.016*mac
0.015*mahasiswa	0.009*ubah	0.012*derajat	0.016*sedia	0.018*kombinasi	0.017*cobit	0.016*byte
63	64	65	66	67	68	69
0.020*minat	0.030*model	0.027*usaha	0.031*tanam	0.025*bahasa	0.063*perangkat	0.053*ajar
0.016*bidang	0.015*nilai	0.023*nilai	0.031*learning	0.022*uji	0.039*perangkat_lunak	0.027*siswa
0.014*langgan	0.012*produksi	0.022*tingkat	0.030*fuzzy	0.019*bal	0.036*lunak	0.017*web
0.014*nilai	0.011*pengaruh	0.017*bisnis	0.028*interface	0.017*nilai	0.028*menu	0.014*sekolah
0.013*uji	0.011*tingkat	0.015*kerja	0.025*cocok	0.016*media	0.026*uji	0.013*guru
0.012*algoritma	0.011*lahan	0.014*ti	0.023*model	0.016*aplikasi	0.024*kontrol	0.013*aplikasi
0.011*kelompok	0.010*skenario	0.014*model	0.022*user	0.015*positif	0.018*bangun	0.012*halaman
0.010*puas	0.010*fokus	0.013*ukur	0.021*konsumsi	0.013*model	0.016*butuh	0.011*tampil
0.009*terminal	0.009*sistem_informasi	0.012*manajemen	0.016*pangan	0.013*tam	0.014*kualitas	0.011*sistem_informasi
0.009*tingkat	0.008*simulasi	0.011*strategi	0.016*tani	0.012*perilaku	0.014*gis	0.011*alumni

70	71	72	73	74	75	76
0.046*media	0.031*lapor	0.052*dosen	0.088*cloud	0.053*nilai	0.038*nilai	0.017*mobile
0.034*social	0.021*usaha	0.029*sistem_ informasi	0.061*bangkit	0.022*rule	0.024*tabel	0.017*pilih
0.027*office	0.016*uang	0.021*giat	0.055*group	0.021*artikel	0.023*fuzzy	0.015*input
0.013*cepat	0.015*akuntansi	0.020*tabel	0.042*grup	0.020*tabel	0.020*uji	0.015*jaring
0.012*reformasi	0.013*kerja	0.019*ajar	0.040*public	0.020*hitung	0.018*metode	0.013*uji
0.011*time	0.012*sistem_ informasi	0.019*varchar	0.036*akses	0.019*bentuk	0.017*rendah	0.013*unit
0.011*aplikasi	0.012*karyawan	0.017*mahasiswa	0.034*skema	0.018*mahasiswa	0.013*hitung	0.012*system
0.011*indonesia	0.012*butuh	0.016*akademik	0.031*token	0.016*basis_data	0.012*pasok	0.012*target
0.011*birokrasi	0.011*kelola	0.016*didik	0.027*enkripsi	0.016*basis	0.012*orang	0.012*model
0.009*dukung	0.011*ukm	0.015*lapor	0.026*identitas	0.015*prediksi	0.011*item	0.012*kabupaten
77	78	79	80	81	82	83
0.047*aplikasi	0.039*internet	0.062*fungsi	0.144*jual	0.059*pesan	0.031*citra	0.034*pasien
0.017*fungsi	0.022*tingkat	0.053*bentuk	0.062*produk	0.042*usaha	0.029*nilai	0.018*periksa
0.016*model	0.019*model	0.045*seederhana	0.043*variabel	0.041*surat	0.027*notasi	0.018*tahap
0.015*database	0.019*listrik	0.042*langkah	0.041*langgan	0.035*bit	0.024*pasang	0.017*ajar
0.015*layan	0.017*pasang	0.040*metode	0.039*node	0.024*tanda	0.022*segmentasi	0.016*gaji
0.015*web	0.017*tahap	0.030*prima	0.033*pasar	0.022*hash	0.016*metode	0.015*admin
0.014*file	0.014*aktivitas	0.028*tabel	0.026*online	0.021*digital	0.016*atur	0.015*staff
0.014*objek	0.013*bumi	0.025*hukum	0.025*media	0.021*dagang	0.016*deteksi	0.014*lapor
0.014*hash	0.013*ajar	0.021*ekspresi	0.022*beli	0.019*simpan	0.016*uji	0.013*sakit
0.012*android	0.012*kapasitas	0.021*variabel	0.020*usia	0.018*tanda_ tangan	0.014*warna	0.013*medis

84	85	86	87	88	89	90
0.073*siswa	0.020*diagram	0.099*giat	0.167*matriks	0.019*ta	0.033*agen	0.112*dokter
0.019*latih	0.016*kerja	0.075*web	0.035*form	0.017*kerja	0.028*input	0.078*pasien
0.018*fuzzy	0.014*tahap	0.062*didik	0.033*kali	0.017*program	0.027*pindah	0.053*sistem_ informasi
0.017*klasifikasi	0.012*sistem_ informasi	0.058*program	0.027*operasi	0.016*monitoring	0.025*skenario	0.052*lapor
0.017*anak	0.012*metode	0.057*laksana	0.025*kolom	0.015*kabupaten	0.023*jual	0.042*kelola
0.017*matriks	0.012*aplikasi	0.047*bantu	0.023*baris	0.015*rumah	0.017*admin	0.037*staf
0.016*iterasi	0.011*lapor	0.040*masyarakat	0.021*algoritma	0.014*teknologi	0.017*form	0.037*periksa
0.016*bit	0.010*koperasi	0.040*fitur	0.016*kali_ matriks	0.013*tani	0.015*implementasi	0.036*temu
0.016*bidang	0.010*case	0.032*kelola	0.015*diagram	0.012*aplikasi	0.015*rule	0.033*janji
0.015*nilai	0.009*butuh	0.031*antarmuka	0.015*tulis	0.011*mahasiswa	0.015*based	0.033*akses
91	92	93	94	95	96	97
0.101*motor	0.037*sistem_ informasi	0.021*dosen	0.051*kelola	0.063*kriteria	0.031*po	0.028*pilih
0.093*parkir	0.024*manfaat	0.015*ruang	0.039*ti	0.051*dosen	0.031*level	0.017*nilai
0.069*mobil	0.020*mudah	0.015*model	0.035*organisasi	0.031*elemen	0.031*processes	0.017*putus
0.042*cm	0.019*web	0.014*kerja	0.034*tata	0.028*banding	0.030*halaman	0.014*kriteria
0.042*low	0.018*terima	0.014*lihat	0.033*tata_kelola	0.028*level	0.030*business	0.013*model
0.035*high	0.014*resep	0.012*butuh	0.029*risiko	0.028*alternatif	0.029*information	0.013*metode
0.033*pin	0.014*information	0.011*capai	0.014*bisnis	0.025*sub	0.029*tanggan	0.012*olah
0.031*sensor	0.013*variabel	0.011*diagram	0.011*tingkat	0.021*guru	0.024*process	0.012*butuh
0.028*kiri	0.012*faktor	0.011*nilai	0.010*nilai	0.020*ahp	0.024*ds	0.011*analisis
0.028*otomatis	0.012*pengaruh	0.011*tampil	0.010*tuju	0.020*mahasiswa	0.020*maturity	0.009*olah_data

98	99
0.036*pattern	0.039*model
0.036*tabel	0.029*bisnis
0.033*support	0.028*proses_bisnis
0.027*web	0.024*aktivitas
0.025*dataset	0.016*usaha
0.025*tampil	0.016*elektronik
0.022*minimal	0.015*material
0.020*pola	0.014*log
0.020*nilai	0.013*process
0.020*berita	0.013*event

B-3. Noun POS Tagged: 150 Topik

0	1	2	3	4	5	6
0.036*aplikasi	0.052*algoritma	0.096*akun	0.087*sop	0.073*siswa	0.061*aplikasi	0.199*resource
0.024*program	0.043*bisnis	0.054*nama	0.022*prosedur	0.059*bidang	0.059*pemilihan	0.093*product
0.024*alumni	0.042*proses_ bisnis	0.051*profil	0.021*matriks	0.057*fuzzy	0.058*warna	0.083*capacity
0.023*pelayanan	0.030*aktivitas	0.048*sampel	0.021*domain	0.052*nilai	0.034*android	0.073*cost
0.023*server	0.028*model	0.045*nilai	0.020*gangguan	0.036*klasifikasi	0.022*suara	0.054*inventory
0.022*komunikasi	0.025*material	0.040*kandidat	0.018*kerja	0.023*pola	0.021*pola	0.049*decision
0.018*pengguna	0.024*log	0.039*sosial	0.017*pelayanan	0.022*algoritma	0.020*foto	0.049*number
0.016*komputer	0.023*nilai	0.036*jaringan	0.015*ukuran	0.020*pelatihan	0.018*pengujian	0.034*unit
0.013*manajemen	0.020*process	0.035*kemiripan	0.014*decision	0.018*kriteria	0.016*variabel	0.033*algorithm
0.013*system	0.019*event	0.030*orang	0.014*staf	0.017*sains	0.014*efek	0.033*research
7	8	9	10	11	12	13
0.066*kualitas	0.044*aplikasi	0.002*aplikasi	0.087*ti	0.045*akreditasi	0.065*alumni	0.112*penilaian
0.054*pernyataan	0.030*kebutuhan	0.002*web	0.028*bisnis	0.034*rumah	0.056*citra	0.080*tenaga
0.052*api	0.023*diagram	0.001*basis	0.026*kelola	0.034*sakit	0.044*segmentasi	0.064*kinerja
0.050*penilaian	0.022*pengguna	0.001*user	0.026*tata_kelola	0.033*rumah_ sakit	0.035*model	0.060*penilaian_ kinerja
0.049*responden	0.017*perangkat	0.001*matriks	0.025*tata	0.031*online	0.031*sekolah	0.051*pendidik
0.034*dimensi	0.017*kesehatan	0.001*kunci	0.020*organisasi	0.030*dokumen	0.025*kandidat	0.025*kerja
0.031*pengujian	0.015*pengemban- gan	0.001*objek	0.019*strategi	0.028*survei	0.019*siswa	0.024*malang
0.030*nomor	0.015*desain	0.001*file	0.019*keselarasan	0.027*rs	0.019*pengujian	0.020*kenaikan
0.028*rumah	0.013*layanan	0.001*server	0.017*adopsi	0.024*penilaian	0.016*pengguna	0.018*unsur
0.027*nilai	0.013*objek	0.001*modifikasi	0.013*universitas	0.022*kertas	0.013*metode	0.018*gaji

14	15	16	17	18	19	20
0.153*rumah	0.061*laporan	0.146*surat	0.040*sms	0.027*ti	0.073*aplikasi	0.152*harga
0.072*pengujian	0.034*akuntansi	0.109*berkas	0.039*biaya	0.019*tingkat	0.037*layanan	0.057*fuzzy
0.060*pengguna	0.029*keuangan	0.035*pembuatan	0.028*tenaga	0.019*kegiatan	0.027*pengembangan	0.048*pembeli
0.055*posisi	0.029*transaksi	0.027*database	0.028*kesehatan	0.017*pratama	0.024*kota	0.046*penjual
0.040*bantuan	0.022*usaha	0.025*laporan	0.024*promosi	0.016*organisasi	0.021*kabupaten	0.028*tingkat
0.039*kerangka	0.022*ukm	0.022*perdagangan	0.023*suku	0.014*bisnis	0.019*lokasi	0.024*air
0.032*kebutuhan	0.019*pengguna	0.019*penelitian	0.021*metode	0.013*kematangan	0.019*pengguna	0.023*variabel
0.027*kota	0.018*aplikasi	0.016*usulan	0.020*media	0.012*layanan	0.018*surabaya	0.023*barang
0.026*basis	0.018*jurnal	0.016*file	0.020*manfaat	0.012*tata	0.017*studi	0.021*transaksi
0.026*keluarga	0.017*penjualan	0.015*usaha	0.018*nilai	0.011*tingkat_kematangan	0.017*biaya	0.021*penurunan
21	22	23	24	25	26	27
0.062*pengguna	0.048*nilai	0.050*barang	0.118*insiden	0.068*layanan	0.047*organisasi	0.041*uji
0.056*dokumen	0.037*peramalan	0.035*form	0.055*aplikasi	0.046*service	0.041*tata	0.039*dokumen
0.037*kegiatan	0.031*perusahaan	0.035*pengujian	0.022*materi	0.038*metode	0.041*kelola	0.036*kalimat
0.035*manajemen	0.030*pegawai	0.031*customer	0.022*pembelajaran	0.035*mysql	0.040*tata_kelola	0.032*diagnosa
0.033*perencanaan	0.027*fuzzy	0.029*cluster	0.019*bentuk	0.031*bisnis	0.019*bisnis	0.031*metode
0.031*anggaran	0.027*metode	0.025*id	0.018*siswa	0.022*identifikasi	0.019*kematangan	0.029*indonesia
0.028*pengendalian	0.023*kriteria	0.025*tabel	0.016*pengguna	0.020*teknologi	0.017*kinerja	0.024*int
0.023*grup	0.023*penilaian	0.017*nama	0.016*evaluasi	0.018*tabel	0.017*tingkat_kematangan	0.023*skenario
0.022*satuan_kerja	0.018*tabel	0.017*harga	0.015*pengembangan	0.018*cluster	0.016*tingkat	0.022*teks
0.022*satuan	0.018*kinerja	0.016*pembelian	0.014*program	0.017*pelayanan	0.013*ti	0.020*positif

28	29	30	31	32	33	34
0.050*dimensi	0.098*internet	0.073*tabel	0.041*aplikasi	0.037*siswa	0.053*web	0.113*kesehatan
0.043*pelanggan	0.036*keuangan	0.069*rancangan	0.031*pertanian	0.036*nilai	0.046*siswa	0.077*fasilitas
0.041*pasien	0.034*penggunaan	0.056*lembaga	0.028*kabupaten	0.029*bali	0.031*pembela- jangan	0.055*pelayanan
0.038*pemerik- saan	0.027*penggunaan _internet	0.047*basis	0.025*program	0.028*bahasa	0.023*bahasa	0.045*rujukan
0.027*tabel	0.026*koperasi	0.045*ujian	0.020*produk	0.025*aplikasi	0.020*xml	0.032*pelayanan _kesehatan
0.018*means	0.024*pelajar	0.034*siswa	0.017*pakar	0.018*anak	0.020*sekolah	0.026*sumber
0.018*type	0.020*tanaman	0.031*rancangan _basis	0.015*tanaman	0.018*sekolah	0.020*kelas	0.025*daya
0.016*jenis	0.018*responden	0.022*kebutuhan	0.015*laporan	0.015*rakyat	0.019*file	0.024*sumber_daya
0.016*telepon	0.017*prestasi	0.022*bisnis	0.012*penjualan	0.014*mata	0.019*aplikasi	0.024*ketersediaan
0.016*fakta	0.015*orang	0.020*kelas	0.012*antarmuka	0.011*halaman	0.018*halaman	0.024*terintegrasi
35	36	37	38	39	40	41
0.094*akses	0.043*algoritma	0.049*model	0.069*enterprise	0.045*budaya	0.077*dosen	0.056*erp
0.050*aktivitas	0.029*anggota	0.022*skenario	0.067*modelan	0.033*video	0.075*kinerja	0.045*proyek
0.035*kondisi	0.023*perhitungan	0.021*kinerja	0.054*pesan	0.028*situs	0.061*penilaian	0.029*perusahaan
0.028*program	0.021*kendaraan	0.020*nilai	0.035*citra	0.028*indonesia	0.058*pengajaran	0.029*penerapan
0.027*formulir	0.019*warna	0.019*pengguna- an	0.022*metode	0.024*pengguna	0.040*penilaian _kinerja	0.029*framework
0.024*pengguna	0.019*fitur	0.015*variabel	0.021*kerangka	0.023*perangkat	0.027*indikator	0.024*implementasi
0.022*kota	0.018*pelajaran	0.012*media	0.021*model	0.019*akun	0.026*evaluasi	0.023*perspektif
0.021*kesenjangan	0.017*kelas	0.012*keuangan	0.020*integrasi	0.019*fitur	0.022*kegiatan	0.023*perubahan
0.019*pegawai	0.017*mata_ pelajaran	0.012*pengem- bangan	0.019*uml	0.019*imple- mentasi	0.021*kehadiran	0.019*organisasi
0.018*pencatatan	0.017*nilai	0.011*anggaran	0.018*fungsi	0.018*guru	0.016*pertemuan	0.019*manajemen

42	43	44	45	46	47	48
0.031*penyakit	0.127*dosen	0.082*game	0.162*aset	0.065*layanan	0.051*dashboard	0.060*media
0.030*gejala	0.092*learning	0.053*virtual	0.121*jaringan	0.065*produksi	0.039*android	0.045*konsumen
0.029*distance	0.086*aplikasi	0.043*wisata	0.057*kondisi	0.053*area	0.021*algoritma	0.044*sosial
0.022*akurasi	0.064*mobile	0.035*objek	0.039*listrik	0.050*anak	0.020*fakultas	0.039*perusahaan
0.021*timur	0.043*pembelajaran	0.033*kriteria	0.036*network	0.038*cakupan	0.020*perusahaan	0.039*media_sosial
0.021*jawa	0.040*aplikasi_mobile	0.029*permainan	0.025*distribusi	0.036*wilayah	0.019*perangkat	0.030*ukm
0.020*attribute	0.037*sisi	0.028*item	0.025*energi	0.035*peramalan	0.019*kriptografi	0.027*proses_bisnis
0.019*lahan	0.036*rancangan	0.024*pembelajaran	0.025*penurunan	0.035*kapasitas	0.017*pengguna	0.026*bisnis
0.018*deteksi	0.032*tampilan	0.022*faktor	0.023*skenario	0.031*sinyal	0.016*layanan	0.022*pelayanan
0.018*system	0.031*perguruan	0.017*online	0.023*average	0.030*propinsi	0.016*mutu	0.021*peta
49	50	51	52	53	54	55
0.041*kerja	0.036*monitoring	0.050*kesehatan	0.045*dosen	0.088*code	0.130*sensor	0.060*pembelajaran
0.032*tenaga	0.023*kamera	0.023*kota	0.028*kegiatan	0.056*citra	0.122*node	0.060*siswa
0.031*tenaga_kerja	0.022*pasien	0.021*ict	0.023*laporan	0.035*matriks	0.049*jaringan	0.054*halaman
0.022*jawa	0.021*atribut	0.018*pengendalian	0.016*tampilan	0.029*audit	0.037*lahan	0.044*media
0.019*bidang	0.019*system	0.017*dinas	0.015*perguruan	0.029*cobit	0.032*pertanian	0.038*media_pembelajaran
0.019*barat	0.019*teknologi	0.017*aplikasi	0.013*keputusan	0.028*ti	0.027*kondisi	0.037*tampilan
0.018*teknologi	0.018*klien	0.015*provinsi	0.012*diagram	0.022*tingkat	0.024*kelembaban	0.036*aplikasi
0.017*provinsi	0.017*internet	0.014*program	0.011*teknologi	0.020*monitor	0.016*energi	0.029*materi
0.016*jawa_barat	0.016*cloud	0.012*kabupaten	0.011*aplikasi	0.018*versi	0.013*tanah	0.028*power
0.016*aktivitas	0.014*produk	0.012*ruang	0.010*akreditasi	0.018*metode	0.013*paket	0.024*listrik

56	57	58	59	60	61	62
0.167*website	0.123*pelanggan	0.045*service	0.043*sehat	0.171*sensor	0.034*risiko	0.096*klasifikasi
0.047*kategori	0.041*produk	0.043*keamanan	0.041*klasifikasi	0.078*obyek	0.021*layanan	0.068*non
0.035*utara	0.031*peramalan	0.035*layanan	0.040*algoritma	0.063*posisi	0.020*service	0.057*daerah
0.035*provinsi	0.030*reko- mendasi	0.024*management	0.032*nilai	0.057*ruangan	0.018*ti	0.045*kondisi
0.031*keamanan	0.028*komponen	0.023*teknologi	0.028*buku	0.048*deteksi	0.018*prosedur	0.036*lahan
0.029*jawa	0.024*pebelian	0.020*audit	0.026*kategori	0.037*gelombang	0.017*til	0.028*metode
0.025*selatan	0.020*harga	0.019*level	0.022*rumah	0.036*database	0.016*pengelolaan	0.022*akurasi
0.024*peringkat	0.017*metode	0.019*kontrol	0.019*dokumen	0.035*area	0.016*akses	0.021*pakar
0.023*celah	0.016*kualitas	0.016*perancangan	0.017*akurasi	0.028*client	0.016*manajemen	0.021*regresi
0.021*sumatera	0.015*nilai	0.015*ti	0.017*metode	0.023*status	0.016*teknologi	0.020*logika
63	64	65	66	67	68	69
0.056*produksi	0.042*pengujian	0.068*dosen	0.042*aplikasi	0.082*media	0.095*layanan	0.056*rumah
0.051*mesin	0.029*gaji	0.042*form	0.029*pengguna	0.056*informa- tion	0.091*share	0.045*sakit
0.046*erp	0.022*halaman	0.029*jadwal	0.022*kebutuhan	0.052*business	0.064*email	0.045*rumah _sakit
0.032*peralatan	0.020*laporan	0.028*ruang	0.019*rute	0.034*process	0.051*aktor	0.033*penggunaan
0.031*kegiatan	0.020*aplikasi	0.017*digital	0.018*usaha	0.031*social	0.034*dosen	0.024*pelayanan
0.031*pemeli- haraan	0.017*imple- mentasi	0.016*laporan	0.015*ti	0.030*level	0.024*unsur	0.021*kesehatan
0.028*aplikasi	0.017*skenario	0.015*diagram	0.013*wisata	0.026*negeri	0.020*permasa- lahan	0.019*teknologi
0.027*dokumen	0.015*user	0.015*pesan	0.013*pariwisata	0.025*organiza- tion	0.019*level	0.015*pasien
0.020*belajar	0.014*php	0.014*pengelolaan	0.012*peta	0.024*maturity	0.018*pelaksana	0.015*fungsi
0.019*risiko	0.014*log	0.013*perkuliahan	0.012*sumatera	0.023*kerja	0.017*wakil	0.013*komunikasi

70	71	72	73	74	75	76
0.029*info	0.039*citra	0.072*perusahaan	0.052*industri	0.111*file	0.079*kompetisi	0.022*pengujian
0.028*dosen	0.023*aplikasi	0.035*perilaku	0.041*pasien	0.065*pesan	0.058*peserta	0.021*perangkat
0.023*mahasiswa	0.021*user	0.035*politik	0.038*kebutuhan	0.060*metode	0.056*jaringan	0.016*tabel
0.020*pencarian	0.018*objek	0.032*keputusan	0.035*query	0.037*enkripsi	0.037*server	0.016*lunak
0.020*titik	0.018*pakar	0.032*budaya	0.032*pengujian	0.032*teks	0.030*komputer	0.015*perangkat _lunak
0.020*nilai	0.017*metode	0.029*pengujian	0.030*diagram	0.029*teknik	0.026*penilaian	0.015*internet
0.019*game	0.015*model	0.027*pengambilan	0.029*kerja	0.027*kunci	0.023*model	0.015*nilai
0.019*jalan	0.015*pengenalan	0.027*pengambilan _keputusan	0.028*waterfall	0.026*aplikasi	0.023*aplikasi	0.013*kategori
0.017*algoritma	0.013*nilai	0.027*laporan	0.027*fuzzy	0.023*isi	0.019*keamanan	0.012*uji
0.016*tujuan	0.012*kebutuhan	0.025*kas	0.026*database	0.023*media	0.018*serangan	0.012*imple- mentasi
77	78	79	80	81	82	83
0.161*sms	0.144*obat	0.056*algoritma	0.031*perangkat	0.073*suara	0.081*cluster	0.047*nilai
0.078*layanan	0.043*jenis	0.036*nilai	0.028*aplikasi	0.034*metode	0.052*mahasiswa	0.037*jaringan
0.076*masyarakat	0.029*metode	0.035*studi	0.027*halaman	0.031*aktivitas	0.043*pusat	0.028*metode
0.058*uji	0.028*ti	0.024*akurasi	0.024*uji	0.030*sampel	0.039*program	0.021*ciri
0.053*bahasa	0.024*klasifikasi	0.022*klasifikasi	0.023*artikel	0.028*sms	0.032*means	0.019*target
0.052*item	0.022*nilai	0.021*penjadwalan	0.022*pengguna	0.023*penghi- tungan	0.031*algoritma	0.019*model
0.036*kecamatan	0.020*percobaan	0.018*aplikasi	0.019*kebutuhan	0.020*perhi- tungan	0.030*titik	0.017*input
0.035*respon	0.018*organisasi	0.018*proyek	0.016*lunak	0.019*kecamatan	0.025*operator	0.017*pengujian
0.031*tingkat	0.016*kompetensi	0.017*mining	0.016*perangkat _lunak	0.018*support	0.025*metode	0.016*algoritma
0.029*kunci	0.016*aplikasi	0.015*mahasiswa	0.015*tabel	0.018*algoritma	0.024*clustering	0.016*error

84	85	86	87	88	89	90
0.068*pemilihan	0.072*listrik	0.116*pembe- lajaran	0.043*pelanggan	0.159*layanan	0.043*metode	0.050*elektronik
0.052*suara	0.053*kapasitas	0.052*mahasiswa	0.039*variabel	0.060*dokumen	0.041*jakarta	0.035*web
0.044*calon	0.038*kunci	0.042*pengem- bangan	0.035*pengguna	0.060*level	0.037*tahap	0.035*tangan
0.040*diagram	0.032*tahap	0.031*learning	0.029*analisis	0.038*service	0.032*air	0.033*dokumen
0.039*pendaftaran	0.032*cluster	0.028*belajar	0.023*tugas	0.037*service_level	0.032*citra	0.033*tanda _tangan
0.036*universitas	0.032*energi	0.025*pemilihan	0.023*nilai	0.026*uraian	0.026*satelit	0.031*tanda
0.035*fakultas	0.027*pesan	0.025*model	0.022*faktor	0.021*penyedia	0.026*kelas	0.024*aplikasi
0.031*permasalahan	0.023*produksi	0.021*mata_ kuliah	0.021*layanan	0.021*penyedia _layanan	0.025*kota	0.016*pembela- jangan
0.029*teknologi	0.023*kelayakan	0.021*komputer	0.018*tabel	0.020*jam	0.025*ruang	0.015*pengem- bangan
0.025*analisa	0.021*protokol	0.021*kuliah	0.017*model	0.020*permintaan	0.024*wilayah	0.013*pemerintah

91	92	93	94	95	96	97
0.043*bahasa	0.059*diagram	0.044*insiden	0.054*penjualan	0.040*biaya	0.103*pengetahuan	0.150*pasangan
0.034*aplikasi	0.034*ciri	0.043*gejala	0.025*dashboard	0.034*negara	0.055*knowledge	0.116*pekerjaan
0.034*indonesia	0.023*case	0.035*nilai	0.019*information	0.030*perangkat	0.055*karyawan	0.069*pemba- ngunan
0.022*bahasa _indonesia	0.021*activity	0.034*penyakit	0.018*perusahaan	0.029*pengem- bangan	0.042*risiko	0.067*proyek
0.019*mahasiswa	0.020*activity _diagram	0.028*tabel	0.017*system	0.027*peneri- maan	0.031*km	0.056*pekerja
0.019*koleksi	0.019*case _diagram	0.026*jenis	0.016*pemesanan	0.026*estimasi	0.029*perusahaan	0.047*rumah
0.016*piranti	0.018*proyek	0.025*prosedur	0.015*pelanggan	0.026*dashboard	0.020*manajemen	0.043*anggaran
0.015*pengetahuan	0.016*laporan	0.025*aturan	0.015*tampilan	0.024*lunak	0.019*identifikasi	0.037*biaya
0.015*uji	0.015*pemodelan	0.023*operasi	0.014*laporan	0.024*perangkat _lunak	0.018*divisi	0.035*aplikasi
0.014*teks	0.015*tahap	0.022*manajemen	0.013*tabel	0.021*aktivitas	0.018*aset	0.029*pintu

98	99	100	101	102	103	104
0.052*promosi	0.126*kapasitas	0.152*obat	0.020*makanan	0.032*permainan	0.072*website	0.040*nilai
0.042*produk	0.072*strategi	0.103*tanaman	0.018*masyarakat	0.026*web	0.039*halaman	0.030*model
0.031*barang	0.070*arus	0.029*struktur	0.016*layanan	0.019*perangkat	0.032*web	0.020*perusahaan
0.029*penjualan	0.057*produksi	0.028*entitas	0.015*login	0.017*aplikasi	0.032*penggunaan	0.018*kebutuhan
0.027*kebutuhan	0.053*barang	0.027*variabel	0.015*jenis	0.016*framework	0.027*teknik	0.018*faktor
0.026*pelanggan	0.052*usaha	0.022*nama	0.014*sosial	0.014*perancangan	0.025*code	0.015*tabel
0.022*ruang	0.042*perencanaan	0.020*aplikasi	0.014*pakar	0.013*pengujian	0.025*authentication	0.013*penggunaan
0.020*toko	0.031*perusahaan	0.020*pengguna	0.014*golongan	0.013*perangkat_lunak	0.023*pengguna	0.012*bisnis
0.017*desain	0.029*menengah	0.018*penyakit	0.012*basis	0.013*lunak	0.018*online	0.012*tingkat
0.016*kisaran	0.024*manfaat	0.016*indonesia	0.012*kegiatan	0.012*php	0.017*serangan	0.012*pengguna
105	106	107	108	109	110	111
0.034*nilai	0.039*usability	0.259*website	0.054*model	0.051*sensor	0.077*diagram	0.076*uji
0.030*fuzzy	0.038*fuzzy	0.242*jurusan	0.045*nilai	0.029*motor	0.039*video	0.040*uji_coba
0.022*algoritma	0.034*model	0.069*rekomendasi	0.029*tabel	0.020*variabel	0.031*obyek	0.040*coba
0.020*himpunan	0.033*rumah	0.055*universitas	0.023*tingkat	0.020*program	0.030*image	0.024*web
0.020*metode	0.027*modul	0.049*pemetaan	0.019*implementasi	0.015*development	0.026*kendaraan	0.024*server
0.018*generasi	0.026*user	0.048*konten	0.017*kriteria	0.014*online	0.024*class	0.023*tabel
0.015*tabel	0.022*logika	0.041*unit	0.017*pengguna	0.013*alat	0.021*activity	0.023*pasangan
0.015*fungsi	0.021*smartphone	0.041*unit_kerja	0.016*hirarki	0.012*teknologi	0.021*case	0.015*modul
0.015*populasi	0.019*logika_fuzzy	0.035*teknik	0.014*dasar	0.012*manusia	0.019*activity_diagram	0.014*algoritma
0.014*kunci	0.019*evaluasi	0.035*kerja	0.013*analisis	0.011*produk	0.019*pengujian	0.014*kendaraan

112	113	114	115	116	117	118
0.028*time	0.089*lahan	0.051*satelit	0.033*perusahaan	0.127*ti	0.126*pengguna	0.067*kapasitas
0.026*ciri	0.058*model	0.047*pengolahan	0.025*bisnis	0.055*analisis	0.055*cloud	0.045*bank
0.024*usability	0.034*produksi	0.035*akuisisi	0.019*tingkat	0.052*strategi	0.047*login	0.041*layanan
0.022*evaluasi	0.033*area	0.034*citra	0.015*tujuan	0.047*cobit	0.047*skema	0.029*capacity
0.018*lokasi	0.031*produk	0.033*perangkat	0.014*kebutuhan	0.045*organisasi	0.036*kunci	0.029*budaya
0.017*pola	0.029*operator	0.024*pengolahan_data	0.012*teknologi	0.036*universitas	0.029*pemilik	0.028*box
0.015*response	0.028*produktivitas	0.021*berkas	0.011*fuzzy	0.031*bisnis	0.028*identitas	0.024*organisasi
0.015*pengguna	0.021*diagram	0.018*kualitas	0.011*sumber_daya	0.029*teknologi	0.027*box	0.023*teknologi
0.015*system	0.015*umpan	0.017*perangkat_lunak	0.011*daya	0.029*teknologi_informasi	0.026*akses	0.018*peningkatan
0.014*blog	0.014*uji	0.017*lunak	0.011*level	0.024*faktor	0.022*kontrol	0.017*ti
119	120	121	122	123	124	125
0.146*matriks	0.055*pengetahuan	0.051*model	0.047*koperasi	0.106*kendaraan	0.224*dashboard	0.088*nilai
0.045*form	0.032*item	0.030*_pembelajaran	0.029*kesehatan	0.047*code	0.106*tampilan	0.029*mata
0.024*operasi	0.029*manajemen	0.028*website	0.022*tingkat	0.037*server	0.106*pengajar	0.028*mahasiswa
0.023*perkalian	0.023*nilai	0.026*online	0.019*laporan	0.027*titik	0.082*int	0.027*mata_kuliah
0.022*kolom	0.022*rekomendasi	0.024*pasar	0.019*dokter	0.026*monitoring	0.067*warehouse	0.027*kuliah
0.021*matriks_matriks	0.019*produk	0.022*indonesia	0.017*keuangan	0.025*keamanan	0.061*mahasiswa	0.024*jaringan
0.021*baris	0.017*variabel	0.022*modal	0.014*skema	0.024*aplikasi	0.057*kuliah	0.023*prediksi
0.020*diagram	0.016*user	0.021*mahasiswa	0.013*pasien	0.021*petugas	0.043*performa	0.021*metode
0.019*algoritma	0.015*kunci	0.021*metode	0.012*metode	0.020*izin	0.037*kejahatan	0.020*basis
0.019*pelanggan	0.014*rating	0.017*penggunaan	0.012*lapisan	0.018*detik	0.037*bentuk	0.020*tabel

126	127	128	129	130	131
0.055*aplikasi	0.036*nilai	0.054*aplikasi	0.030*aplikasi	0.072*pembelajaran	0.078*input
0.034*file	0.026*aplikasi	0.035*perusahaan	0.019*user	0.050*media	0.033*mobil
0.033*pengguna	0.022*model	0.032*android	0.017*arsitektur	0.045*siswa	0.028*harian
0.018*algoritma	0.015*tabel	0.032*erp	0.015*jaringan	0.044*aplikasi	0.022*form
0.013*fungsi	0.012*analisis	0.032*uang	0.014*server	0.035*media _pembelajaran	0.022*penjualan
0.013*tahap	0.012*diagram	0.029*modul	0.014*kode	0.022*variabel	0.022*output
0.012*pengujian	0.012*pengguna	0.025*ujian	0.014*pengguna	0.022*belajar	0.020*admin
0.012*enkripsi	0.011*pengujian	0.022*metode	0.013*keamanan	0.019*sql	0.018*produksi
0.011*lingkungan	0.011*metode	0.019*versi	0.012*service	0.019*nilai	0.017*high
0.011*java	0.011*perpustakaan	0.018*open	0.011*fungsi	0.018*perintah	0.017*halaman
132	133	134	135	136	137
0.028*metode	0.043*kinerja	0.031*dimensi	0.069*kualitas	0.046*diagram	0.071*kriteria
0.011*pengolahan	0.042*pengukuran	0.030*aplikasi	0.055*perangkat	0.021*paket	0.040*keputusan
0.011*analisis	0.031*faktor	0.022*teknologi	0.035*lunak	0.020*produksi	0.039*alternatif
0.011*model	0.028*pengukuran _kinerja	0.022*tik	0.035*perangkat _lunak	0.018*laporan	0.020*nilai
0.010*laporan	0.027*nilai	0.021*kota	0.032*kebutuhan	0.017*pokok	0.019*pendukung
0.010*pengolahan _data	0.022*kota	0.019*adaptasi	0.023*kontrol	0.016*jasa	0.016*pendukung _keputusan
0.009*keputusan	0.021*pemerintah	0.017*pengembangan	0.020*verifikasi	0.014*rancangan	0.016*metode
0.009*mahasiswa	0.019*sukses	0.017*teknologi _informasi	0.019*penjaminan	0.013*aplikasi	0.015*elemen
0.009*kebutuhan	0.015*kriteria	0.016*pemanfaatan	0.019*nilai	0.011*testing	0.015*perbandingan
0.009*tahap	0.015*daerah	0.014*organisasi	0.019*penerapan	0.010*uml	0.013*ahp

138	139	140	141	142	143
0.036*pemilihan	0.031*peta	0.031*dokumen	0.023*siswa	0.056*sim	0.043*model
0.025*tahapan	0.029*item	0.021*layanan	0.021*aplikasi	0.025*digital	0.042*metode
0.025*tahap	0.028*analisa	0.015*analisis	0.020*kebutuhan	0.024*bukti	0.038*tabel
0.024*motor	0.019*strategi	0.015*nilai	0.015*prioritas	0.022*web	0.037*fungsi
0.019*sensor	0.018*nilai	0.013*aplikasi	0.014*proyek	0.021*file	0.034*bentuk
0.015*elektronik	0.016*tabel	0.013*bisnis	0.012*nilai	0.021*minat	0.032*nilai
0.014*metode	0.016*perusahaan	0.013*ti	0.012*online	0.021*model	0.023*langkah
0.014*low	0.016*mining	0.011*tabel	0.012*fungsi	0.017*kelompok	0.016*parameter
0.014*rangkaian	0.016*kondisi	0.010*organisasi	0.010*ujian	0.015*telepon	0.016*prima
0.013*tes	0.016*teknologi	0.009*metode	0.009*perangkat	0.015*nilai	0.016*variabel
144	145	146	147	148	149
0.274*kinerja	0.093*support	0.088*investasi	0.045*nilai	0.052*kegiatan	0.063*nilai
0.146*penilaian	0.091*pattern	0.081*resiko	0.040*metode	0.046*knowledge	0.038*hash
0.135*karyawan	0.059*web	0.064*tingkat	0.037*kriteria	0.038*materi	0.032*fungsi
0.121*penilaian_kinerja	0.058*perpustakaan	0.051*komposisi	0.028*keputusan	0.036*fitur	0.029*database
0.044*performance	0.056*berita	0.049*pengembalian	0.022*penilaian	0.029*pengetahuan	0.026*skema
0.040*perusahaan	0.050*pola	0.046*rekomendasi	0.020*bobot	0.023*model	0.025*tabel
0.018*ukur	0.033*accuracy	0.033*komponen	0.020*total	0.022*pembelajaran	0.025*algoritma
0.017*penulis	0.024*enterprise	0.032*pengguna	0.019*derajat	0.021*benda	0.023*support
0.017*alat	0.022*architecture	0.031*form	0.019*tahap	0.021*penanda	0.023*fungsi_hash
0.014*sdm	0.021*mobile	0.031*fungsi	0.017*wawancara	0.019*media	0.020*pelanggan

LAMPIRAN C

C-1. Top Topics: Tanpa Stemming 50 Topik

Topik36		Topik34		Topik5		Topik17		Topik37	
0.01531	tata	0.03292	layanan	0.01116	pengguna	0.02208	model	0.01256	pemilihan
0.01383	tata_kelola	0.01292	kebutuhan	0.01037	nilai	0.01113	nilai	0.01147	metode
0.01351	kelola	0.01080	analisis	0.00924	user	0.00929	tabel	0.00842	dokumen
0.01320	dashboard	0.01041	bisnis	0.00840	aktivitas	0.00733	sistem_ informasi	0.00769	produksi
0.01305	bisnis	0.00913	ti	0.00790	analisis	0.00709	pelanggan	0.00758	pendaftaran
0.01275	ti	0.00864	service	0.00710	metode	0.00665	rumah	0.00753	suara
0.01106	organisasi	0.00850	kondisi	0.00675	system	0.00643	skenario	0.00721	diagram
0.01037	kebutuhan	0.00777	teknologi	0.00628	memiliki	0.00630	kondisi	0.00694	aplikasi
0.00959	teknologi	0.00735	tabel	0.00613	teknologi	0.00625	listrik	0.00672	information
0.00861	proses_bisnis	0.00706	sop	0.00580	perangkat	0.00621	user	0.00638	matriks
0.00813	perusahaan	0.00677	aktivitas	0.00577	sop	0.00609	metode	0.00610	memiliki
0.00716	insiden	0.00633	prioritas	0.00561	situs	0.00606	penggunaan	0.00585	tingkat
0.00707	fungsi	0.00621	metode	0.00515	pelanggan	0.00566	produk	0.00554	diperoleh
0.00706	sumber	0.00605	manajemen	0.00512	halaman	0.00561	memiliki	0.00551	nilai
0.00699	manajemen	0.00589	aplikasi	0.00500	salah	0.00531	jaringan	0.00546	basis
0.00679	teknologi_ informasi	0.00563	penilaian	0.00498	tabel	0.00509	variabel	0.00536	level
0.00646	level	0.00562	teknik	0.00487	admin	0.00495	konsumen	0.00525	business
0.00644	aplikasi	0.00560	nilai	0.00480	pengujian	0.00479	pengguna	0.00515	process
0.00642	tujuan	0.00557	level	0.00464	penggunaan	0.00451	tingkat	0.00511	processes
0.00619	pengelolaan	0.00552	dokumen	0.00461	pengembangan	0.00450	system	0.00511	tabel
-0.61271		-0.75074		-0.79455		-0.87037		-1.04047	

Topik12		Topik13		Topik15		Topik6		Topik44	
0.02561	siswa	0.02269	erp	0.03286	pengguna	0.03061	aplikasi	0.04800	sistem_ informasi
0.01491	aplikasi	0.01884	perusahaan	0.02329	variabel	0.01607	pengguna	0.04053	kesehatan
0.01303	dokumen	0.01534	kriteria	0.02260	nilai	0.01500	kebutuhan	0.02904	pelayanan
0.01144	nilai	0.01074	aplikasi	0.01942	faktor	0.00949	game	0.02658	web
0.01009	tabel	0.01066	biaya	0.01794	penggunaan	0.00850	perangkat	0.02529	penjualan
0.00781	memiliki	0.01034	bisnis	0.01216	model	0.00756	user	0.02336	fasilitas
0.00752	sekolah	0.01021	alternatif	0.01131	indikator	0.00712	fungsi	0.01601	rujukan
0.00679	tahap	0.00968	sistem_ informasi	0.00916	analisis	0.00709	permainan	0.01152	transaksi
0.00667	guru	0.00900	proyek	0.00906	hipotesis	0.00699	pengujian	0.01109	database
0.00660	indonesia	0.00873	enterprise	0.00887	layanan	0.00639	virtual	0.01071	pelayanan_ kesehatan
0.00633	belajar	0.00769	penilaian	0.00846	penerapan	0.00609	fungsional	0.01069	ketersediaan
0.00621	analisis	0.00737	pengembangan	0.00838	positif	0.00597	objek	0.01059	dokter
0.00615	dimensi	0.00732	adaptasi	0.00812	perceived	0.00578	web	0.01034	terintegrasi
0.00613	user	0.00708	berfokus	0.00800	adopsi	0.00562	server	0.00978	sumber_ daya
0.00592	file	0.00646	tabel	0.00790	tabel	0.00554	fitur	0.00950	daya
0.00565	kode	0.00644	proses_ bisnis	0.00785	berpengaruh	0.00509	code	0.00945	sumber
0.00546	pembelajaran	0.00625	penerapan	0.00767	penerimaan	0.00508	sesuai	0.00862	layanan
0.00530	kelas	0.00619	karyawan	0.00721	pengaruh	0.00499	algoritma	0.00817	rumah_ sakit
0.00510	implementasi	0.00592	manajemen	0.00698	signifikan	0.00497	pengembangan	0.00789	integrasi
0.00492	nasional	0.00577	modul	0.00684	memiliki	0.00481	file	0.00774	laporan
-1.09020		-1.10019		-1.10071		-1.12254		-1.21239	

C-2. Top Topics: Tanpa Stemming 100 Topik

Topik7		Topik1		Topik64		Topik36		Topik93	
0.01719	matriks	0.01637	jaringan	0.02004	aplikasi	0.04736	layanan	0.04083	nilai
0.01006	sistem_ informasi	0.01208	teknologi	0.01261	pengguna	0.01526	akses	0.02676	metode
0.00969	form	0.01120	learning	0.01135	laporan	0.01473	service	0.01908	fuzzy
0.00919	diagram	0.01007	perguruan	0.01126	monitoring	0.01235	dokumen	0.01296	keputusan
0.00902	pelanggan	0.00925	user	0.00915	information	0.01198	sop	0.01182	kriteria
0.00832	pengujian	0.00840	perangkat	0.00887	system	0.01183	aplikasi	0.01165	tabel
0.00746	metode	0.00814	mahasiswa	0.00828	basis	0.01070	pengguna	0.00983	uji
0.00727	pengolahan	0.00799	mobile	0.00800	server	0.01057	kondisi	0.00975	variabel
0.00660	keputusan	0.00794	keputusan	0.00764	process	0.00900	aktivitas	0.00962	model
0.00657	laporan	0.00741	pembelajaran	0.00758	basis_data	0.00879	analisis	0.00868	coba
0.00635	halaman	0.00715	keamanan	0.00737	sesuai	0.00851	manajemen	0.00852	algoritma
0.00617	jadwal	0.00696	orang	0.00707	level	0.00827	management	0.00784	aturan
0.00586	pengolahan_ data	0.00671	perancangan	0.00647	business	0.00817	pembuatan	0.00749	himpunan
0.00558	perangkat	0.00663	memiliki	0.00643	processes	0.00788	itil	0.00707	analisis
0.00529	menu	0.00650	arsitektur	0.00590	bentuk	0.00771	terkait	0.00696	rendah
0.00528	pengembangan	0.00632	lunak	0.00589	kebutuhan	0.00770	standar	0.00680	menentukan
0.00508	perancangan	0.00602	perangkat_ lunak	0.00577	tabel	0.00690	desk	0.00673	perhitungan
0.00504	kriteria	0.00591	program	0.00572	barang	0.00663	prosedur	0.00648	akurasi
0.00503	desain	0.00557	internet	0.00571	user	0.00621	tabel	0.00644	tingkat
0.00491	model	0.00551	kondisi	0.00570	pembuatan	0.00590	sesuai	0.00624	pelanggan
-0.83817		-0.84614		-0.86714		-0.87368		-0.94622	

Topik62		Topik10		Topik71		Topik91		Topik20	
0.03297	model	0.01710	kinerja	0.05382	ti	0.03298	aplikasi	0.02806	diagram
0.01147	nilai	0.01499	level	0.05196	organisasi	0.01488	kebutuhan	0.01406	sistem_ informasi
0.00966	metode	0.01263	anggaran	0.01766	kelola	0.01251	layanan_ publik	0.01044	kebutuhan
0.00941	kebutuhan	0.01241	kegiatan	0.01624	tata_kelola	0.01156	layanan	0.01002	tahap
0.00932	algoritma	0.01206	tujuan	0.01619	tata	0.01143	publik	0.00956	case
0.00921	aset	0.01183	audit	0.01126	faktor	0.01002	pengembangan	0.00840	kegiatan
0.00921	aktivitas	0.01084	keamanan	0.01112	teknologi	0.00842	nilai	0.00744	model
0.00850	organisasi	0.00794	bisnis	0.01078	pratama	0.00761	memiliki	0.00727	pengguna
0.00806	bisnis	0.00764	unit	0.01063	budaya	0.00742	siswa	0.00694	alumni
0.00727	analisis	0.00747	manajemen	0.00971	tujuan	0.00671	metode	0.00673	perancangan
0.00688	distribusi	0.00703	sistem_ informasi	0.00894	manajemen	0.00665	mata	0.00667	dokumen
0.00638	tabel	0.00703	system	0.00832	penggunaan	0.00660	diagram	0.00659	desain
0.00602	kualitas	0.00697	keuangan	0.00763	kerja	0.00607	web	0.00655	analisis
0.00595	pendekatan	0.00674	berbasis	0.00757	bisnis	0.00593	pembangunan	0.00612	bisnis
0.00581	pelanggan	0.00651	organisasi	0.00736	pengelolaan	0.00563	model	0.00599	laporan
0.00572	dihasilkan	0.00628	laporan	0.00711	pelayanan	0.00550	perangkat	0.00585	aplikasi
0.00572	tingkat	0.00625	indikator	0.00685	governance	0.00540	tabel	0.00581	activity
0.00526	listrik	0.00592	manfaat	0.00680	manfaat	0.00493	sesuai	0.00568	uml
0.00511	berbasis	0.00590	universitas	0.00662	penerapan	0.00493	service	0.00562	evaluasi
0.00501	kinerja	0.00584	performance	0.00642	kematangan	0.00491	desain	0.00542	database
-0.95922		-0.97677		-0.99211		-1.02323		-1.03393	

C-3. Top Topics: Tanpa Stemming 150 Topik

Topik122		Topik74		Topik127		Topik60		Topik37	
0.02068	pengguna	0.01431	kerja	0.01534	aplikasi	0.09224	erp	0.02554	layanan
0.01882	perangkat	0.01398	tata	0.01063	sistem_ informasi	0.04910	perusahaan	0.02103	metode
0.01659	lunak	0.01378	tata_kelola	0.00918	metode	0.04466	modul	0.01726	server
0.01422	perangkat_ lunak	0.01328	kelola	0.00830	kebutuhan	0.02843	bisnis	0.01481	aplikasi
0.01353	sesuai	0.01270	karyawan	0.00710	pengolahan	0.02721	karyawan	0.01335	identifikasi
0.01280	rumah	0.01212	kebutuhan	0.00687	kriteria	0.02673	open	0.01253	system
0.01219	aplikasi	0.01143	teknologi	0.00662	pengembangan	0.02258	proses_bisnis	0.01203	bisnis
0.01202	implementasi	0.01102	sumber	0.00599	komputer	0.01649	penerapan	0.01138	user
0.01108	kebutuhan	0.01100	absensi	0.00585	pengolahan_ data	0.01640	aplikasi	0.00915	service
0.00884	indonesia	0.01012	daya	0.00585	tahap	0.01530	user	0.00833	publik
0.00852	office	0.01001	teknologi_ informasi	0.00580	biaya	0.01494	sistem_ informasi	0.00777	pendekatan
0.00845	pengujian	0.00967	sumber_daya	0.00569	manajemen	0.01311	resource	0.00769	teknologi
0.00837	fitur	0.00915	perusahaan	0.00568	model	0.01287	kegiatan	0.00748	tabel
0.00831	fungsional	0.00762	android	0.00523	alternatif	0.01161	sumber	0.00726	monitoring
0.00809	analisis	0.00742	kinerja	0.00500	teknologi	0.01120	sumber_daya	0.00620	tujuan
0.00807	pembuatan	0.00738	pengembangan	0.00472	medan	0.01085	daya	0.00609	analisis
0.00792	halaman	0.00738	tujuan	0.00471	diagram	0.00934	planning	0.00609	arsitektur
0.00737	memiliki	0.00730	aplikasi	0.00462	system	0.00908	simulasi	0.00607	tahap
0.00728	user	0.00719	berbasis	0.00462	permasalahan	0.00880	sesuai	0.00598	pengguna
0.00723	tabel	0.00686	kegiatan	0.00461	menghasilkan	0.00873	model	0.00563	pelayanan_ publik
-0.70033		-0.80748		-0.83185		-0.85588		-0.88459	

Topik59		Topik11		Topik52		Topik6		Topik144	
0.05733	aplikasi	0.04958	siswa	0.03366	perusahaan	0.05529	tata	0.12449	ti
0.03399	layanan	0.03657	pembelajaran	0.02706	dokumen	0.04847	kelola	0.04564	organisasi
0.01371	diagram	0.01853	media	0.02548	ti	0.04767	tata_kelola	0.04368	kelola
0.01341	web	0.01833	belajar	0.02062	nilai	0.03993	organisasi	0.04334	tata
0.01205	berbasis	0.01624	learning	0.01898	analisis	0.01354	tingkat	0.04111	tata_kelola
0.01165	user	0.01489	nilai	0.01779	keputusan	0.01316	kematangan	0.02277	pratama
0.01155	pengguna	0.01479	media_ pembelajaran	0.01359	tingkat	0.01113	universitas	0.01491	tujuan
0.01115	kota	0.01373	item	0.01323	perilaku	0.01079	bisnis	0.01293	bisnis
0.00980	kesehatan	0.01248	materi	0.01095	kinerja	0.01009	tingkat_ kematangan	0.01246	governance
0.00973	desain	0.01117	sekolah	0.01064	politik	0.01003	budaya	0.01194	cobit
0.00969	pengem- bangan	0.01084	kelas	0.01002	tabel	0.00957	manajemen	0.01051	pengelolaan
0.00841	admin	0.00850	berbasis	0.00953	memiliki	0.00949	ti	0.01017	kematangan
0.00833	kebutuhan	0.00810	tampilan	0.00930	organisasi	0.00829	pengelolaan	0.00959	manajemen
0.00609	teknologi	0.00786	aplikasi	0.00926	pengambilan	0.00813	teknologi	0.00928	strategi
0.00605	uml	0.00741	tabel	0.00899	budaya	0.00760	terkait	0.00891	pusat
0.00589	case	0.00687	ujian	0.00885	kategori	0.00760	perguruan	0.00865	pelayanan
0.00562	perancangan	0.00676	sma	0.00861	variabel	0.00672	tujuan	0.00863	penerapan
0.00558	analisis	0.00674	jenis	0.00848	manajemen	0.00662	keputusan	0.00841	tingkat_ kematangan
0.00539	mobile	0.00665	memiliki	0.00768	pengujian	0.00646	governance	0.00780	kinerja
0.00530	studi	0.00657	tahap	0.00761	bisnis	0.00608	kuesioner	0.00737	sesuai
-0.89001		-0.89662		-0.90057		-0.92795		-0.93437	

C-4. Top Topics: Stemming 50 Topik

Topik 43		Topik 1		Topik 11		Topik 38		Topik 4	
0.01630	ti	0.02655	layan	0.03906	sistem_informasi	0.06124	uji	0.01407	organisasi
0.01520	kelola	0.01559	service	0.01618	nilai	0.01490	halaman	0.01255	tingkat
0.01518	butuh	0.01536	akses	0.01394	uji	0.01364	aplikasi	0.01226	butuh
0.01442	budaya	0.01189	aplikasi	0.01304	tujuan	0.01284	perangkat	0.01226	tujuan
0.01405	kerja	0.01137	management	0.01261	kualitas	0.01281	tabel	0.01186	portofolio
0.01320	teknologi	0.01136	dokumen	0.01145	nyata	0.01253	perangkat_lunak	0.01083	manajemen
0.01143	aplikasi	0.00949	analisis	0.01120	aplikasi	0.01227	lunak	0.01041	sistem_informasi
0.01003	organisasi	0.00942	sop	0.01088	jual	0.01220	bangun	0.01014	model
0.00949	tata	0.00917	sesuai	0.00983	usaha	0.01126	web	0.01004	insiden
0.00795	perangkat	0.00906	kerja	0.00848	tampil	0.01124	implementasi	0.00992	prosedur
0.00768	sesuai	0.00893	kondisi	0.00841	lihat	0.01099	lihat	0.00918	olah
0.00729	tata_kelola	0.00890	aktivitas	0.00836	knowledge	0.00923	basis	0.00906	kerja
0.00727	implementasi	0.00858	teknologi	0.00807	salah	0.00910	lapor	0.00862	dukungan
0.00678	teknologi_informasi	0.00839	risiko	0.00792	responden	0.00892	basis_data	0.00860	metode
0.00668	fungsi	0.00834	tangan	0.00782	nomor	0.00832	metode	0.00741	perintah
0.00648	sistem_informasi	0.00831	kait	0.00736	input	0.00801	terima	0.00727	giat
0.00604	tujuan	0.00828	manajemen	0.00716	mudah	0.00800	gaji	0.00724	terapi
0.00596	nilai	0.00748	butuh	0.00641	masuk	0.00724	rancang	0.00706	bisnis
0.00579	usaha	0.00743	tahap	0.00634	implementasi	0.00710	butuh	0.00687	tahap
0.00578	jalan	0.00733	tingkat	0.00628	tingkat	0.00698	bahasa	0.00675	olah_data
-0.47162		-0.56008		-0.58669		-0.72126		-0.77748	

Topik 13		Topik 49		Topik 20		Topik 24		Topik 14	
0.01605	usaha	0.01544	password	0.02326	butuh	0.01990	kerja	0.03486	model
0.01450	form	0.01333	login	0.02132	kualitas	0.01618	kriteria	0.01833	kapasitas
0.01350	tampil	0.01211	aplikasi	0.01959	media	0.01610	usaha	0.01755	skenario
0.01335	jual	0.01108	peta	0.01864	perangkat	0.01319	putus	0.01613	nilai
0.01321	produk	0.01001	orang	0.01617	layan	0.01177	tingkat	0.01369	variabel
0.01140	pasar	0.00986	sistem_ informasi	0.01416	usaha	0.01165	nilai	0.01296	simulasi
0.01124	butuh	0.00898	user	0.01377	surat	0.01012	sehat	0.01236	listrik
0.01101	diagram	0.00840	tahap	0.01351	sesuai	0.00931	ambil	0.01061	layan
0.01052	karyawan	0.00827	lokasi	0.01332	tingkat	0.00841	dukung	0.00938	pengaruh
0.01017	langgan	0.00817	simpan	0.01192	aplikasi	0.00809	banding	0.00916	time
0.00922	promosi	0.00770	anggota	0.01142	lunak	0.00781	layan	0.00810	kondisi
0.00841	mudah	0.00746	rumah	0.01132	perangkat_ lunak	0.00751	faktor	0.00806	produksi
0.00790	web	0.00738	based	0.01049	sosial	0.00724	sakit	0.00785	pasang
0.00741	ancang	0.00723	bentuk	0.01034	berkas	0.00699	ukur	0.00782	banding
0.00726	dukung	0.00720	kota	0.00998	sedia	0.00687	adopsi	0.00766	lihat
0.00693	website	0.00713	masuk	0.00959	teknologi	0.00686	alternatif	0.00751	tahap
0.00661	objek	0.00654	komponen	0.00940	manfaat	0.00650	tabel	0.00743	tingkat
0.00637	desain	0.00650	tuju	0.00918	main	0.00610	sedia	0.00729	panas
0.00635	layan	0.00641	mobile	0.00914	fasilitas	0.00596	pasien	0.00714	capai
0.00624	database	0.00632	identifikasi	0.00897	manajemen	0.00580	aplikasi	0.00705	kerja
-0.84836		-0.85193		-0.87135		-0.88584		-0.89212	

C-5. Top Topics: Stemming 100 Topik

Topik 40		Topik 39		Topik 34		Topik 97		Topik 13	
0.02893	layan	0.02408	uji	0.01434	nilai	0.02798	pilih	0.02985	framework
0.01148	ti	0.01754	perangkat	0.01164	learning	0.01705	nilai	0.01802	fungsi
0.01136	teknologi	0.01685	aplikasi	0.01148	aman	0.01678	putus	0.01403	file
0.01097	kerja	0.01428	perangkat_ lunak	0.01142	aktivitas	0.01446	kriteria	0.01389	aplikasi
0.01071	service	0.01426	lunak	0.01126	butuh	0.01336	model	0.01354	identifikasi
0.01000	aplikasi	0.01407	butuh	0.00897	model	0.01304	metode	0.01258	layan
0.00999	tingkat	0.01306	tampil	0.00878	fuzzy	0.01230	olah	0.01188	web
0.00954	kondisi	0.01082	implementasi	0.00810	tuju	0.01167	butuh	0.01054	lihat
0.00932	management	0.01015	diagram	0.00764	amal	0.01056	analisis	0.01034	ajar
0.00926	akses	0.00995	lihat	0.00748	faktor	0.00941	olah_data	0.00992	ancang
0.00920	kelola	0.00975	sesuai	0.00744	metode	0.00896	tingkat	0.00951	konsep
0.00904	nilai	0.00970	user	0.00742	ajar	0.00880	dukung	0.00937	analisis
0.00886	sop	0.00958	lapor	0.00731	perangkat	0.00810	ambil	0.00862	model
0.00875	kait	0.00929	desain	0.00703	tingkat	0.00737	berkas	0.00793	bangun
0.00863	manajemen	0.00841	mudah	0.00695	pilih	0.00720	surat	0.00782	akses
0.00854	butuh	0.00836	dashboard	0.00691	tahap	0.00694	tahap	0.00778	standar
0.00848	tahap	0.00805	analisis	0.00679	lihat	0.00688	ambil_putus	0.00753	dkk
0.00841	sesuai	0.00800	bangun	0.00641	kerja	0.00669	sesuai	0.00739	kode
0.00836	dokumen	0.00796	web	0.00627	usaha	0.00656	bentuk	0.00732	nama
0.00823	analisis	0.00793	model	0.00615	akses	0.00648	variabel	0.00714	service
-0.52207		-0.62282		-0.71735		-0.72508		-0.75639	

Topik 41		Topik 1		Topik 48		Topik 85		Topik 47	
0.04938	aplikasi	0.07212	nilai	0.02698	layan	0.02044	diagram	0.09052	ti
0.02917	layan	0.02441	kerja	0.01897	model	0.01568	kerja	0.03729	cobit
0.02349	publik	0.02294	kriteria	0.01637	usaha	0.01382	tahap	0.03412	soft
0.02339	layan_ publik	0.02279	metode	0.01287	butuh	0.01231	sistem_ informasi	0.03300	organisasi
0.02098	butuh	0.01808	bobot	0.01281	metode	0.01200	metode	0.03040	hard
0.01312	bangun	0.01453	prioritas	0.01172	tingkat	0.01155	aplikasi	0.02947	level
0.01263	kota	0.01439	putus	0.00965	bisnis	0.01144	lapor	0.01832	nilai
0.01112	surabaya	0.01276	hitung	0.00931	pengaruh	0.01024	koperasi	0.01775	kompetensi
0.00847	sedia	0.01255	butuh	0.00924	nilai	0.01013	case	0.01742	analisis
0.00825	server	0.01082	dukung	0.00903	putus	0.00938	butuh	0.01684	teknologi
0.00768	manfaat	0.01081	banding	0.00892	bit	0.00933	uji	0.01542	daya
0.00761	teknologi	0.01034	ahp	0.00886	box	0.00859	dokumen	0.01525	teknologi_ informasi
0.00736	model	0.00977	alternatif	0.00832	ambil	0.00809	ajar	0.01471	sdm
0.00690	biaya	0.00955	teknik	0.00827	tabel	0.00748	hitung	0.01463	sumber_daya
0.00653	user	0.00941	total	0.00824	tuju	0.00736	tingkat	0.01433	rekomendasi
0.00649	tenaga	0.00919	daftar	0.00797	implementasi	0.00734	uang	0.01395	tahap
0.00612	web	0.00878	tahap	0.00782	kondisi	0.00720	tulis	0.01392	sumber
0.00611	mobile	0.00739	tabel	0.00741	ti	0.00659	rancang	0.01275	meta
0.00567	sesuai	0.00721	tingkat	0.00725	sesuai	0.00653	jalan	0.01212	bisnis
0.00543	tuju	0.00666	derajat	0.00720	teknologi	0.00624	sehat	0.01182	kuesioner
-0.78161		-0.80856		-0.82022		-0.82593		-0.83051	

C-6. Top Topics: Stemming 150 Topik

Topik 123		Topik 35		Topik 72		Topik 69		Topik 45	
0.01373	server	0.03357	ti	0.04273	nilai	0.02168	management	0.02011	lihat
0.01339	aplikasi	0.02203	usaha	0.02196	metode	0.01668	teknologi	0.01346	nilai
0.01139	uji	0.01591	teknologi	0.01919	kerja	0.01618	layan	0.01293	user
0.01106	komputer	0.01535	kerja	0.01913	tabel	0.01419	aktivitas	0.01172	analisis
0.01058	jaring	0.01509	tingkat	0.01854	uji	0.01308	kelola	0.01155	tabel
0.01053	software	0.01456	nilai	0.01178	model	0.01267	tahap	0.01148	cari
0.01043	database	0.01338	organisasi	0.01121	faktor	0.01168	tingkat	0.01127	butuh
0.00924	layan	0.01257	bisnis	0.01062	hitung	0.01058	rancang	0.00990	aplikasi
0.00913	program	0.01093	pengaruh	0.01033	usaha	0.01052	problem	0.00983	share
0.00849	tingkat	0.00959	analisis	0.01012	tahap	0.01040	teknologi_ informasi	0.00931	rancang
0.00792	cepat	0.00939	sesuai	0.00992	amal	0.01006	daya	0.00897	perangkat
0.00764	tabel	0.00939	implementasi	0.00965	lihat	0.01001	level	0.00891	salah
0.00761	mudah	0.00922	uji	0.00946	salah	0.00999	butuh	0.00888	sesuai
0.00757	kode	0.00886	sumber	0.00892	ambil	0.00967	karyawan	0.00841	kerja
0.00717	metode	0.00851	butuh	0.00852	sesuai	0.00962	kondisi	0.00835	input
0.00663	lihat	0.00785	perilaku	0.00799	butuh	0.00951	analisis	0.00801	tahap
0.00663	nilai	0.00779	model	0.00798	kelompok	0.00936	sumber_daya	0.00780	kuliah
0.00660	kerja	0.00775	kerangka	0.00730	rendah	0.00908	tuju	0.00759	fungsi
0.00659	butuh	0.00750	teknologi_ informasi	0.00723	hubung	0.00879	sumber	0.00752	bentuk
0.00657	web	0.00740	tahap	0.00713	tampil	0.00859	sistem_ informasi	0.00728	tuju
-0.41023		-0.51924		-0.54988		-0.59781		-0.60796	

Topik 86		Topik 19		Topik 117		Topik 140		Topik 131	
0.03448	aplikasi	0.02089	sistem_ informasi	0.02425	kualitas	0.02183	faktor	0.03664	diagram
0.02820	web	0.01733	kerja	0.02134	perangkat	0.02127	manfaat	0.01559	uji
0.02198	butuh	0.01598	ajar	0.02055	aktivitas	0.02120	usaha	0.01364	butuh
0.01602	tampil	0.01415	olah	0.01993	butuh	0.01886	erp	0.01313	kerja
0.01367	sistem_ informasi	0.01396	butuh	0.01702	kontrol	0.01819	modul	0.01261	dosen
0.01184	fungsi	0.01254	lapor	0.01540	lunak	0.01484	pengaruh	0.01168	case
0.01134	fitur	0.01219	metode	0.01522	aplikasi	0.01452	terap	0.01148	tahap
0.01107	guru	0.01196	tahap	0.01518	perangkat_ lunak	0.01434	kualitas	0.01142	sistem_ informasi
0.00995	ruang	0.01113	olah_data	0.01308	verifikasi	0.01401	open	0.01119	lapor
0.00952	bangun	0.01073	giat	0.01199	kelola	0.01373	karyawan	0.01036	catat
0.00774	objek	0.01053	laksana	0.01179	prosedur	0.01261	aplikasi	0.01012	activity
0.00767	ajar	0.01046	rumah	0.01176	sesuai	0.01247	butuh	0.00965	rancang
0.00728	manfaat	0.00907	desain	0.01173	standar	0.01239	terima	0.00934	transaksi
0.00686	salah	0.00864	medan	0.01050	akses	0.01206	sistem_ informasi	0.00913	atur
0.00630	application	0.00857	rancang	0.01024	iso	0.01113	manusia	0.00827	model
0.00621	model	0.00807	tuju	0.00970	kait	0.01080	user	0.00806	desain
0.00611	analisis	0.00748	putus	0.00947	jamin	0.01052	uji	0.00802	ancang
0.00593	teknologi	0.00736	terap	0.00885	tabel	0.00985	human	0.00777	analisis
0.00580	lihat	0.00728	aplikasi	0.00844	manajemen	0.00953	model	0.00759	class
0.00576	kota	0.00726	system	0.00767	analisis	0.00953	teknologi	0.00748	tampil
-0.64169		-0.69653		-0.72582		-0.75054		-0.77896	

C-7. Top Topics: Noun POS Tagged 50 Topik

Topik 18		Topik 26		Topik 45		Topik 49		Topik 36	
0.01997	layanan	0.04554	perusahaan	0.03272	tata	0.03327	perusahaan	0.03552	aplikasi
0.01870	teknologi	0.02290	bisnis	0.03189	kelola	0.01741	pengetahuan	0.02297	pengguna
0.01320	aplikasi	0.02228	penjualan	0.03126	tata_kelola	0.01652	karyawan	0.01682	kebutuhan
0.01118	pengguna	0.02206	risiko	0.02770	organisasi	0.01590	nilai	0.01663	perangkat
0.01027	tabel	0.01632	manajemen	0.01497	aplikasi	0.01525	bisnis	0.01622	web
0.00965	penggunaan	0.01560	teknologi	0.01492	web	0.01147	variabel	0.01520	diagram
0.00962	faktor	0.01323	enterprise	0.01364	model	0.01114	produk	0.01228	user
0.00925	kerja	0.01148	ti	0.01234	diagram	0.01084	proses_bisnis	0.01190	pengujian
0.00830	kebutuhan	0.01115	organisasi	0.01136	bidang	0.01024	metode	0.01121	rumah
0.00730	organisasi	0.00984	keputusan	0.01111	pendaftaran	0.01019	analisis	0.01065	implementasi
0.00693	sop	0.00946	kerangka	0.01070	nilai	0.00958	knowledge	0.00928	analisis
0.00656	pelayanan	0.00938	kondisi	0.00992	manajemen	0.00884	tabel	0.00878	lunak
0.00646	pengembangan	0.00923	teknologi_informasi	0.00987	tingkat	0.00812	jaringan	0.00877	perangkat_lunak
0.00637	teknologi_informasi	0.00908	information	0.00953	kegiatan	0.00807	penjualan	0.00813	fitur
0.00631	dokumen	0.00904	system	0.00953	minat	0.00798	input	0.00786	adopsi
0.00628	model	0.00890	database	0.00941	program	0.00788	model	0.00760	model
0.00626	media	0.00864	tujuan	0.00930	framework	0.00761	laporan	0.00748	objek
0.00603	kondisi	0.00816	aplikasi	0.00911	fungsi	0.00749	aplikasi	0.00736	android
0.00592	pemerintah	0.00779	bank	0.00849	bisnis	0.00741	diagram	0.00699	pengembangan
0.00591	masyarakat	0.00766	analisa	0.00832	pengelolaan	0.00706	customer	0.00597	teknologi
-0.82849		-0.83548		-0.88460		-0.89832		-0.91575	

Topik 42		Topik 35		Topik 38		Topik 23		Topik 28	
0.04702	nilai	0.01919	resiko	0.05146	pengguna	0.10291	ti	0.03220	layanan
0.03114	kriteria	0.01543	nilai	0.02940	nilai	0.03870	organisasi	0.02682	ti
0.02590	pelanggan	0.01526	video	0.01522	tingkat	0.03545	bisnis	0.02136	service
0.02517	metode	0.01460	kebutuhan	0.01316	evaluasi	0.02719	faktor	0.01359	analisis
0.02026	keputusan	0.01328	website	0.01289	aplikasi	0.02316	strategi	0.01146	manajemen
0.01805	fuzzy	0.01259	tahap	0.01276	model	0.01976	keselarasan	0.01089	tingkat
0.01757	alternatif	0.01250	model	0.01145	responden	0.01887	cobit	0.01032	akses
0.01727	bobot	0.01221	metode	0.01072	game	0.01658	nilai	0.01004	kesehatan
0.01337	rekomendasi	0.01201	akun	0.01064	tabel	0.01493	model	0.00980	management
0.01280	tabel	0.01146	tingkat	0.01001	pengujian	0.01347	xyz	0.00977	teknologi
0.01259	kinerja	0.01112	penilaian	0.00981	virtual	0.01246	universitas	0.00932	prosedur
0.01169	produk	0.01103	pengguna	0.00908	penggunaan	0.01023	tingkat	0.00912	nilai
0.01113	model	0.01055	kegiatan	0.00847	perangkat	0.00920	level	0.00898	itil
0.01074	variabel	0.01052	perencanaan	0.00844	metode	0.00871	kerja	0.00832	tugas
0.00971	ahp	0.01050	manajemen	0.00815	login	0.00816	manajemen	0.00783	pengelolaan
0.00911	analisis	0.01032	budaya	0.00802	pembelajaran	0.00802	pengukuran	0.00779	aktivitas
0.00886	penilaian	0.01028	aktivitas	0.00793	tahap	0.00779	rekomendasi	0.00758	kelola
0.00821	matriks	0.00976	analisis	0.00775	teknologi	0.00764	budaya	0.00746	tata_kelola
0.00809	perhitungan	0.00890	desain	0.00714	mahasiswa	0.00753	penilaian	0.00739	tata
0.00803	tingkat	0.00845	halaman	0.00656	web	0.00727	faktor_faktor	0.00709	pelayanan
-0.94080		-0.95388		-0.96368		-1.14486		-1.15421	

C-8. Top Topics: Noun POS Tagged 100 Topik

Topik 52		Topik 96		Topik 87		Topik 99		Topik 9	
0.03030	aplikasi	0.03788	aplikasi	0.02515	model	0.06064	ti	0.06287	nilai
0.01769	sensor	0.01985	kebutuhan	0.02172	analisis	0.02873	perusahaan	0.02002	pemilihan
0.01668	tabel	0.01708	perangkat	0.02056	pembelajaran	0.02501	layanan	0.01602	tingkat
0.01334	nilai	0.01310	pengembangan	0.01728	aplikasi	0.02034	organisasi	0.01464	analisis
0.01273	server	0.01263	metode	0.01524	nilai	0.01848	manajemen	0.01334	tabel
0.01015	implementasi	0.01196	diagram	0.01229	aktivitas	0.01557	kinerja	0.01261	kriteria
0.00983	database	0.01161	web	0.01105	bisnis	0.01509	kerangka	0.01223	rekomendasi
0.00920	jaringan	0.01136	android	0.00982	siswa	0.01499	teknologi	0.01176	metode
0.00894	kondisi	0.01116	bahasa	0.00979	tabel	0.01372	analisis	0.01094	alternatif
0.00862	komunikasi	0.01097	perancangan	0.00909	pengem- bangan	0.01351	keputusan	0.01078	model
0.00766	pengem- bangan	0.00969	pengguna	0.00879	materi	0.01306	faktor	0.01018	pengujian
0.00744	layanan	0.00953	teknologi	0.00838	web	0.01297	teknologi_ informasi	0.00974	mahasiswa
0.00739	manajemen	0.00942	fungsi	0.00824	uji	0.01273	bank	0.00964	solusi
0.00738	penerapan	0.00863	tabel	0.00814	pengguna	0.01240	bisnis	0.00934	keputusan
0.00722	user	0.00841	lunak	0.00774	metode	0.01235	adopsi	0.00903	perhitungan
0.00720	kebutuhan	0.00837	perangkat_ lunak	0.00772	teknologi	0.01100	daya sumber_ daya	0.00872	algoritma
0.00717	kriteria	0.00819	file	0.00769	media	0.01007	daya	0.00857	system
0.00701	teknologi	0.00796	information	0.00757	kegiatan	0.00990	sumber	0.00780	pengguna
0.00665	model	0.00747	objek	0.00735	online	0.00980	kondisi	0.00773	perancangan
0.00642	perangkat	0.00711	indonesia	0.00727	desain	0.00917	nilai	0.00720	input
-0.65366		-0.72800		-0.75151		-0.79081		-0.80724	

Topik 55		Topik 75		Topik 63		Topik 86		Topik 45	
0.06233	pengetahuan	0.05853	tata	0.06750	layanan	0.05953	faktor	0.02845	kegiatan
0.02787	knowledge	0.05783	kelola	0.03617	sop	0.04127	kota	0.02581	anggaran
0.02766	km	0.05693	tata_kelola	0.03020	service	0.02795	organisasi	0.02417	analisis
0.01998	karyawan	0.04992	ti	0.01589	aplikasi	0.02544	teknologi	0.02412	perencanaan
0.01987	perusahaan	0.04651	organisasi	0.01426	akses	0.02371	pemerintah	0.02144	laporan
0.01853	manajemen	0.02044	bisnis	0.01311	surabaya	0.02135	variabel	0.01885	universitas
0.01654	strategi	0.01966	kematangan	0.01303	kondisi	0.01634	kuesioner	0.01801	keuangan
0.01523	organisasi	0.01891	tingkat	0.01268	pelayanan	0.01633	daerah	0.01629	penilaian
0.01438	aplikasi	0.01724	manajemen	0.01221	dokumen	0.01615	pengguna	0.01417	akuntansi
0.01232	management	0.01617	tingkat_kem atangan	0.01220	uji	0.01605	manfaat	0.01399	strategi
0.01202	bisnis	0.01542	tujuan	0.01193	itil	0.01539	penerimaan	0.01375	cobit
0.01055	dampak	0.01281	cobit	0.01188	prosedur	0.01491	implement- tasi	0.01208	tujuan
0.00880	perguruan	0.01249	teknologi	0.01009	kota	0.01476	penerapan	0.01107	kebutuhan
0.00873	nilai	0.01080	teknologi_ informasi	0.00973	standar	0.01423	sukses	0.01043	tim
0.00854	analisis	0.01073	pengelolaan	0.00970	management	0.01363	kualitas	0.01020	pelaksanaan
0.00818	tabel	0.01050	strategi	0.00935	permintaan	0.01249	tingkat	0.00988	organisasi
0.00801	divisi	0.00907	kebijakan	0.00858	pembuatan	0.01216	positif	0.00984	kinerja
0.00793	variabel	0.00850	kuesioner	0.00845	pengguna	0.01200	responden	0.00959	pengguna
0.00780	system	0.00744	budaya	0.00824	aktivitas	0.01141	pengem- bangan	0.00948	bisnis
0.00758	metode	0.00730	universitas	0.00780	tabel	0.01112	kemudahan	0.00943	desain
-0.85926		-0.92189		-0.92494		-0.96587		-0.98958	

C-9. Top Topics: Noun POS Tagged 150 Topik

Topik 8		Topik 80		Topik 140		Topik 129		Topik 104	
0.04399	aplikasi	0.03109	perangkat	0.03125	dokumen	0.02973	aplikasi	0.04012	nilai
0.02960	kebutuhan	0.02817	aplikasi	0.02086	layanan	0.01852	user	0.03012	model
0.02336	diagram	0.02720	halaman	0.01525	analisis	0.01747	arsitektur	0.02025	perusahaan
0.02232	pengguna	0.02397	uji	0.01511	nilai	0.01454	jaringan	0.01843	kebutuhan
0.01721	perangkat	0.02280	artikel	0.01338	aplikasi	0.01410	server	0.01795	faktor
0.01665	kesehatan	0.02240	pengguna	0.01319	bisnis	0.01385	kode	0.01465	tabel
0.01542	pengem- bangan	0.01889	kebutuhan	0.01266	ti	0.01381	pengguna	0.01295	penggunaan
0.01479	desain	0.01617	lunak	0.01078	tabel	0.01343	keamanan	0.01243	bisnis
0.01312	layanan	0.01605	perangkat_ lunak	0.00968	organisasi	0.01242	service	0.01237	tingkat
0.01252	objek	0.01524	tabel	0.00909	metode	0.01141	fungsi	0.01234	pengguna
0.01177	web	0.01467	pengujian	0.00889	fungsi	0.01108	web	0.01220	analisis
0.01149	user	0.01437	tombol	0.00815	strategi	0.01106	database	0.01098	variabel
0.01090	analisis	0.01402	pelanggan	0.00780	kebutuhan	0.01047	perangkat	0.01090	metode
0.01035	metode	0.01278	information	0.00760	tahap	0.01043	system	0.01036	mahasiswa
0.01009	software	0.01119	indonesia	0.00756	kerja	0.00940	pengujian	0.00996	teknik
0.00990	penilaian	0.01119	daftar	0.00747	service	0.00913	layanan	0.00983	resiko
0.00966	laporan	0.01077	metode	0.00688	sop	0.00846	metode	0.00940	manajemen
0.00933	tahap	0.01049	bahasa	0.00686	perusahaan	0.00835	transaksi	0.00929	hubungan
0.00920	tampilan	0.01049	implementasi	0.00679	manajemen	0.00834	teknologi	0.00850	layanan
0.00907	masyarakat	0.01024	diagram	0.00651	pembuatan	0.00801	android	0.00845	kondisi
-0.73989		-0.76017		-0.76839		-0.78591		-0.78841	

Topik 61		Topik 94		Topik 92		Topik 71		Topik 127	
0.03433	risiko	0.05416	penjualan	0.05884	diagram	0.03950	citra	0.03625	nilai
0.02089	layanan	0.02530	dashboard	0.03421	ciri	0.02277	aplikasi	0.02638	aplikasi
0.02013	service	0.01924	information	0.02301	case	0.02132	user	0.02195	model
0.01814	ti	0.01763	perusahaan	0.02097	activity	0.01844	objek	0.01489	tabel
0.01781	prosedur	0.01743	system	0.02042	activity_ diagram	0.01769	pakar	0.01214	analisis
0.01724	itil	0.01639	pemesanan	0.01873	case_diagram	0.01699	metode	0.01199	diagram
0.01644	pengelolaan	0.01492	pelanggan	0.01809	proyek	0.01522	model	0.01150	pengguna
0.01629	akses	0.01452	tampilan	0.01583	laporan	0.01473	pengenalan	0.01107	pengujian
0.01618	manajemen	0.01430	laporan	0.01494	pemodelan	0.01312	nilai	0.01096	metode
0.01556	teknologi	0.01285	tabel	0.01460	tahap	0.01190	kebutuhan	0.01090	perpustakaan
0.01481	sop	0.01259	produk	0.01444	kebutuhan	0.01051	input	0.01022	uji
0.01465	management	0.01192	keputusan	0.01347	analisis	0.01017	program	0.00909	coba
0.01413	kondisi	0.01174	konsumen	0.01318	perancangan	0.00984	pengguna	0.00869	kegiatan
0.01371	aktivitas	0.01160	wilayah	0.01315	administrasi	0.00967	file	0.00865	uji_coba
0.01367	kontrol	0.01109	form	0.01268	blok	0.00927	teknologi	0.00847	teknologi
0.01360	pengguna	0.01069	metode	0.01198	dokumen	0.00913	interface	0.00809	perangkat
0.01295	standar	0.01036	input	0.01197	pembuatan	0.00848	form	0.00808	alat
0.01154	iso	0.01004	cabang	0.01174	desain	0.00818	vektor	0.00776	bentuk
0.01119	aset	0.00969	bentuk	0.01157	penulis	0.00805	diagram	0.00765	android
0.01110	teknologi_ informasi	0.00956	kebutuhan	0.01156	pengujian	0.00746	system	0.00717	keluaran
-0.93983		-0.97597		-0.98764		-1.02195		-1.03207	

LAMPIRAN D

D-1. Daftar Dokumen Pengujian

Index	Nama File	Judul	Author	Keyword
1	106	Analisa Kesuksesan Sistem Informasi Akademik (SIKAD) di Perguruan Tinggi dengan menggunakan D & M IS Success Model (Studi Kasus: ITS Surabaya)	Ardhini Warih Utami, Febriyian Samopa	D & M IS Success, sistem informasi akademik, pemodelan persamaan struktural
2	121	Prototipe <i>Data Warehouse</i> Perikanan Tangkap dan Budidaya	Rully Agus Hendrawan	data warehouse, business intelligence, fishery
3	146	Perencanaan Strategis Teknologi Informasi Pada Sektor Publik menggunakan <i>Kerangka The Open Group Architecture Framework (TOGAF)</i>	Mario Glendi Kasenda, Eko Nugroho, Selo Sulistyono	Rencana Strategi, Sistem Informasi, TOGAF ADM
4	16	Studi Kelayakan Investasi Perluasan Jaringan TV Kabel Pada PT. XYZ dengan Metode <i>Cost & Benefit Analysis</i>	Tony Dwi Susanto, Aditya Zulkifli, Anisah Herdiyanti	Payback Period, Net Present Value (NPV), Return On Investment (ROI), Profitability index (PI), Internal Rate of Return (IRR) dan Average Rate of Return (ARR)
5	186	Analisis Pola Asosiasi dan Sekuensial Data Rekam Medis RSUD Dr. H. Slamet Martodirdjo Pamekasan dengan Teknik <i>Data Mining</i> menggunakan Algoritma Apriori	Nilam Ramadhani, Badar Said	penggalan data, data rekam medis, association rule, sequential pattern, apriori

6	200	Perancangan Desain Ruang <i>Data Center</i> menggunakan Standar TIA-942 (Studi Kasus: Puslitbang Jalan dan Jembatan)	Dimas Sigit Dewandaru, Arief Bachtiar	Data center, Desain, TIA-942
7	206	Evaluasi Keamanan Informasi menggunakan Indeks Keamanan Informasi pada Kantor Wilayah Ditjen Perbendaharaan Negara Jawa Timur	Mustaqim Siga, Tony Dwi Susanto, Beki Cahyo Hidayanto	evaluasi, keamanan informasi, indeks KAMI
8	213	Perancangan Sistem Pakar Minat Mahasiswa berdasarkan Kurikulum Program Studi Teknik Informatika Universitas XYZ	Budi Arifitama	Sistem pakar, perminatan TI
9	226	Purwarupa Aplikasi Pembelajaran <i>SQL</i> Interaktif berbasis Web dengan Penerapan <i>Gamification</i>	Radityo Prasentianto Wibowo, Fachri Hilmi Romdhoni	aplikasi interaktif, gamification, pembelajaran <i>SQL</i>
10	270	Implementasi Metode <i>Incremental</i> dalam membangun Aplikasi <i>Use Case Point</i> Pada Perusahaan DTS	Mukhamad Faiz Fanani, Sholiq, Feby Artwodini Muqtadiroh	Estimasi harga perangkat, Effort, UCP, Incremental Model, SKPL, DPPL
11	285	Segmentasi dan Evaluasi Loyalitas Pelanggan Distributor Produk Etikal Farmasi berdasarkan Nilai Pelanggan	Rully Agus Hendrawan, Amalia Utamima, Annisa Husna	pengelolaan hubungan pelanggan, segmentasi pelanggan, model RFM
12	298	Pemanfaatan Konten Pembelajaran bagi Siswa Sekolah Menengah Kejuruan	Nugroho Nurcahyono, Adhistya Erna Permanasari	<i>e-learning</i> , pembelajaran, SMK
13	306	Perancangan Sistem Informasi Penentuan Rute Pengiriman Barang dengan Metode	Arief S. Gunawan, Evasaria M. Sipayung, Agustinus Wiguno	pengiriman, penentuan rute, Ant Colony Optimization (ACO)

		<i>Ant Colony Optimization</i> Studi Kasus: PT. XYZ		
14	312	Perancangan KPI Individu untuk Target Pencapaian Keamanan Informasi Studi Kasus Kementerian XYZ	Arie Kusumawati, Nuraini Purwandari	COBIT 5, ITIL, KPI (Key Performance Indicator), ITSM, Keamanan data dan informasi
15	352	Eksperimen Sistem Klasifikasi Analisa Sentimen <i>Twitter</i> pada Akun Resmi Pemerintah Kota Surabaya berbasis Pembelajaran Mesin	Nuke Y. A. Faradhillah, Renny P. Kusumawardani, Irmasari Hafidz	Klasifikasi Teks, Twitter, Algoritma Naïve Bayes, Algoritma Support Vector Machine (SVM)
16	373	Sejarah, Tantangan, dan Faktor Keberhasilan dalam Pengembangan <i>E-Learning</i>	Riska Agustina, Paulus Insap Santosa, Ridi Ferdiana	e-learning, tantangan, faktor sukses
17	393	Analisis dan Desain Sistem <i>Monitoring</i> Aktivitas Pekerja Sosial Masyarakat	Arif Wibisono, Amna Shifia Nisafani, Rully Agus Hendrawan	Pekerja Sosial Lapangan, Analisis dan Desain Sistem, GIS
18	40	Penerapan <i>Multiple Stock</i> dalam Sistem Informasi Persediaan Bahan pada Resto Bambo Jakarta	Soleh, Sudi Herman, Dwi Puguh Kurniawan	persediaan, bahan baku, sistem informasi, stok, multiple stock units
19	420	Implementasi Algoritma Prim dengan Teori <i>Graph</i> pada <i>WPF Graph</i>	Trinanda Syahputra, Dedi Setiawan	Algoritma Prim, Graph, Struktur Data, TSP
20	91	Implementasi Algoritma <i>Fuzzy-Madm</i> dalam Menentukan Pola Tanaman Pangan Kabupaten Jayapura, Papua	Fred Melvery Degei, Andeka Rocky Tanaamah, Alz Danny Wowor	Fuzzy Multi-Attribute Decision Making (FMADM)

D-2. Hasil Pengujian *Stemming* 50 Topik – *Minimal Author Size=1*

Model *Stemming* 50 Topik: Dokumen 1

	Author	Score	Size
88	Annis Paramita Dilla	0.603198	1
130	Bayu Kelana	0.599306	1
261	Febrianto	0.597287	1
362	Irna Yuniar	0.597284	1
701	Wan Maritul Kifti	0.597046	1
280	Ger Mustofa	0.597024	1
506	Nurmaini Dalimunthe	0.596988	1
167	Deni Arifianto	0.596848	1
600	Sejati Waluyo	0.596503	1
374	Joko Lianto Buliali	0.596332	2

Model *Stemming* 50 Topik: Dokumen 2

	Author	Score	Size
623	Soetjipto	0.596465	1
125	Balqis Lembah Mahersmi	0.594037	1
119	Aullia Rachman	0.593517	1
699	Wahjoe Witjaksono	0.593176	1
500	Nur Abidin	0.592887	1
254	Falahah Suprpto	0.592823	1
562	Rini Juliana Sipahutar	0.592735	1
100	Arie Yandi Saputra	0.592442	1
204	Dwi Pratama	0.591774	1
434	Mesran	0.591719	1

Model *Stemming* 50 Topik: Dokumen 3

	Author	Score	Size
143	Budi Yuwono	0.609110	1
21	Agnes Djohan	0.605734	1
399	Lanto Ningrayati Amali	0.605538	1
32	Agustinus Fritz Wijaya	0.602771	1
426	Marcel	0.602679	2
387	Khairunnisa Rofifah	0.602409	1
396	Lailatul Fitriana R	0.601377	1
296	Hanung Nindito Prasetyo	0.600956	2
117	Augie David Manuputty	0.600897	1
692	Vianny Utami Tjhin	0.600340	1

Model Stemming 50 Topik: Dokumen 4

	Author	Score	Size
397	Lailil Muflikhah	0.601245	1
472	Nadia Roosmalita Sari	0.601205	1
420	M. Tanzil Furqon	0.600441	1
697	W. Lisa Yunita	0.599933	1
191	Diyurman Gea	0.598293	1
374	Joko Lianto Buliali	0.597948	2
654	Sylvia Lavinia	0.597890	1
282	Gita Indah Marthasari	0.597758	1
634	Stefanus	0.597708	1
726	Wiwik Anggraeni	0.597509	5

Model Stemming 50 Topik: Dokumen 5

	Author	Score	Size
551	Ridha Sefina Samosir	0.597459	1
397	Lailil Muflikhah	0.597156	1
558	Rin Rin Meilani Salim	0.596944	1
191	Diyurman Gea	0.596413	1
420	M. Tanzil Furqon	0.596388	1
697	W. Lisa Yunita	0.595947	1
282	Gita Indah Marthasari	0.595891	1
490	Noorman Rinanto	0.595565	1
200	Dwi Cahya Astriya Nugraha	0.595499	1
722	Windi Eka Y. R	0.595137	1

Model Stemming 50 Topik: Dokumen 6

	Author	Score	Size
589	Safrina Amini	0.592378	1
336	I Made Gede Sri Artha	0.592227	1
623	Soetjipto	0.592015	1
125	Balqis Lembah Mahersmi	0.591899	1
36	Ahmad Fitriansyah	0.591787	1
43	Albert Sagala	0.591706	1
561	Rinanza Z. Alhamri	0.591553	1
188	Dimas Sigit Dewandaru	0.591535	1
708	Widyo Nugroho	0.591284	1
580	Rosa Lestari Fardani	0.591138	1

Model Stemming 50 Topik: Dokumen 7

	Author	Score	Size
143	Budi Yuwono	0.596435	1
362	Irna Yuniar	0.594832	1
551	Ridha Sefina Samosir	0.594491	1
558	Rin Rin Meilani Salim	0.594382	1
565	Risky Dinal Ardianto	0.594300	1
692	Vianny Utami Tjhin	0.594012	1
654	Sylvia Lavinia	0.593978	1
674	Tony	0.593954	1
32	Agustinus Fritz Wijaya	0.593857	1
62	Anak Agung Gde Agung	0.593772	1

Model Stemming 50 Topik: Dokumen 8

	Author	Score	Size
539	Ramen Antonov Purba	0.596323	1
310	Hendra Dinata	0.596266	1
630	Sri Widodo	0.595147	1
620	Siti Muslimah Widyastuti	0.595036	1
658	Tanti Kristanti	0.594886	1
629	Sri Tomo	0.594819	1
394	Kustanto	0.594682	1
562	Rini Juliana Sipahutar	0.594572	1
556	Riki Andri Yusda	0.594440	1
668	Theopilus Sitepu	0.594142	1

Model Stemming 50 Topik: Dokumen 9

	Author	Score	Size
623	Soetjipto	0.599496	1
523	Purba Daru Kusuma	0.597400	1
100	Arie Yandi Saputra	0.596370	1
456	Muhammad Amin	0.596036	1
464	Muhammad Sabir Ramadhan	0.595763	1
254	Falahah Suprpto	0.595749	1
125	Balqis Lembah Mahersmi	0.595723	1
434	Mesran	0.595671	1
119	Aullia Rachman	0.595453	1
754	Yuri Prihantono	0.595452	1

Model Stemming 50 Topik: Dokumen 10

	Author	Score	Size
22	Agung Brastama Putra	0.598674	1
	730 Yanuarini N.S	0.595912	1
360	Irlando Moggi Prakoso	0.594854	1
	36 Ahmad Fitriansyah	0.594848	1
	43 Albert Sagala	0.594829	1
188	Dimas Sigit Dewandaru	0.594813	1
	397 Lailil Muflikhah	0.594650	1
	697 W. Lisa Yunita	0.594638	1
	377 Joko Suprianto	0.594590	1
38	Ahmad Mukhlason	0.594549	1

Model Stemming 50 Topik: Dokumen 11

	Author	Score	Size
472	Nadia Roosmalita Sari	0.600579	1
	191 Diyurman Gea	0.598112	1
	397 Lailil Muflikhah	0.597998	1
282	Gita Indah Marthasari	0.597612	1
	420 M. Tanzil Furqon	0.597227	1
	654 Sylvia Lavinia	0.597011	1
726	Wiwik Anggraeni	0.596887	5
	634 Stefanus	0.596817	1
443	Mohammad Fajar	0.596654	2
	697 W. Lisa Yunita	0.596624	1

Model Stemming 50 Topik: Dokumen 12

	Author	Score	Size
	246 Fajar Annas	0.600651	1
	486 Nisaul Barokati	0.599936	1
427	Mardiyanto Wiyogo	0.599503	1
	626 Sri Hermuningsih	0.599478	1
	545 Restu Amalia	0.599379	1
	623 Soetjipto	0.599328	1
501	Nur Aini Rakhmawati	0.599246	1
	655 Syukur Ikhsani	0.599102	1
	390 Kristi Wardani	0.599040	1
	17 Afdhal Syafnur	0.598297	1

Model Stemming 50 Topik: Dokumen 13

	Author	Score	Size
200	Dwi Cahya Astriya Nugraha	0.594666	1
282	Gita Indah Marthasari	0.594370	1
191	Diyurman Gea	0.594267	1
472	Nadia Roosmalita Sari	0.593631	1
207	Dyah Erny Herwindiati	0.593611	2
159	Dede Sutarya	0.593598	1
397	Lailil Muflikhah	0.593328	1
0	A.A. Gede Bagus Ariana	0.593194	1
318	Herman Saputra	0.592888	1
527	Putu Agung Bayupati	0.592867	1

Model Stemming 50 Topik: Dokumen 14

	Author	Score	Size
143	Budi Yuwono	0.598268	1
692	Viany Utami Tjhin	0.595952	1
426	Marcel	0.595752	2
387	Khairunnisa Rofifah	0.595467	1
32	Agustinus Fritz Wijaya	0.595430	1
362	Irna Yuniar	0.594939	1
294	Hanim Maria Astuti	0.594722	5
296	Hanung Nindito Prasetyo	0.594662	2
643	Sulistiyowati	0.594134	1
565	Risky Dinal Ardianto	0.594104	1

Model Stemming 50 Topik: Dokumen 15

	Author	Score	Size
529	Rachmadita Andreswari	0.599972	1
623	Soetjipto	0.599287	1
88	Annis Paramita Dilla	0.597387	1
343	Ifrina Nuritha	0.596567	1
501	Nur Aini Rakhmawati	0.596374	1
585	Ryan Randy Suryono	0.596161	1
261	Febrianto	0.596135	1
701	Wan Maritul Kifti	0.595353	1
363	Irwan Haryo Yudananto	0.595343	1
505	Nurizal Dwi Priandani	0.595293	1

Model Stemming 50 Topik: Dokumen 16

	Author	Score	Size
626	Sri Hermuningsih	0.600753	1
390	Kristi Wardani	0.600306	1
5	Achmad Solichin	0.599608	1
17	Afdhal Syafnur	0.599573	1
477	Neni Mulyani	0.599424	3
110	Arridha Zikra Syah	0.599326	3
571	Rizky Tahara Shita	0.599286	1
246	Fajar Annas	0.599262	1
82	Angraini	0.599256	1
249	Fajara Kurniawan N.H	0.599240	1

Model Stemming 50 Topik: Dokumen 17

	Author	Score	Size
143	Budi Yuwono	0.596411	1
343	Ifrina Nuritha	0.594598	1
21	Agnes Djohan	0.594284	1
124	Bahar	0.594049	2
119	Aullia Rachman	0.593906	1
223	Elvira Azany	0.593881	1
268	Fitriyadi	0.593851	1
640	Sugiyanto	0.593634	1
525	Purwanti	0.593527	1
526	Putri Ayu Rahayu	0.593502	1

Model Stemming 50 Topik: Dokumen 18

	Author	Score	Size
472	Nadia Roosmalita Sari	0.596223	1
200	Dwi Cahya Astriya Nugraha	0.593910	1
565	Risky Dinal Ardianto	0.593415	1
397	Lailil Muflikhah	0.593183	1
207	Dyah Erny Herwindiati	0.593141	2
562	Rini Juliana Sipahutar	0.592887	1
0	A.A. Gede Bagus Ariana	0.592514	1
318	Herman Saputra	0.592478	1
654	Sylvia Lavinia	0.592471	1
420	M. Tanzil Furqon	0.592441	1

Model Stemming 50 Topik: Dokumen 19

	Author	Score	Size
529	Rachmadita Andreswari	0.595921	1
351	Imelda Atastina	0.595855	1
357	Indra Setia Rahmat	0.595816	1
343	Ifrina Nuritha	0.595757	1
79	Angelina Prima K	0.594817	1
166	Dedy Suwandi	0.594812	1
363	Irwan Haryo Yudananto	0.594175	1
56	Amalia Utamima	0.594042	2
407	Lintang Kesumastuti	0.593905	1
301	Harsih Rianto	0.593867	1

Model Stemming 50 Topik: Dokumen 20

	Author	Score	Size
472	Nadia Roosmalita Sari	0.599198	1
191	Diyurman Gea	0.598841	1
277	Gayatri Dwi Santika	0.598512	1
282	Gita Indah Marthasari	0.597368	1
742	Yohannes Kurniawan	0.597225	1
726	Wiwik Anggraeni	0.596972	5
684	Ufi Sanjaya	0.596892	1
181	Dian Ramadhani	0.596280	1
443	Mohammad Fajar	0.596210	2
722	Windi Eka Y. R	0.596112	1

D-3. Hasil Pengujian *Stemming* 50 Topik – *Minimal Author Size=3*

Model *Stemming* 50 Topik: Dokumen 1

	Author	Score	Size
34	Rully Agus Hendrawan	0.594052	5
	14 Erma Suryani	0.593864	8
	36 Sofu Risqi Y.S.	0.593072	3
	13 Ely Rosely	0.592983	3
23	Mahendrawathi ER	0.592886	3
	45 Zulfi Azhar	0.589878	3
0	Alz Danny Wowor	0.586109	4
	35 Safrian Aswati	0.585816	8
	37 Stanley Karouw	0.585693	3
	27 Nisfu Asrul Sani	0.585642	3

Model *Stemming* 50 Topik: Dokumen 2

	Author	Score	Size
	37 Stanley Karouw	0.591422	3
	28 Oleh Soleh	0.590877	3
33	Royyana M. Ijtihadie	0.589921	3
	32 Rida Indah Fariani	0.588466	3
	24 Mira Ziveria	0.586825	3
	27 Nisfu Asrul Sani	0.586390	3
42	William Ramdhan	0.586369	3
	12 Desy Iba Ricoida	0.586266	3
	7 Arif Wibisono	0.586030	3
3	Andre Parvian Aristio	0.585940	3

Model *Stemming* 50 Topik: Dokumen 3

	Author	Score	Size
20	Hanim Maria Astuti	0.599362	5
	37 Stanley Karouw	0.594896	3
	38 Titus Kristanto	0.594796	4
	30 Raras Tyasnurita	0.594454	3
	28 Oleh Soleh	0.594409	3
17	Feby Artwodini Muqtadiroh	0.594351	8
	18 Fransiskus Adikara	0.594349	4
29	Radityo Prasetyanto Wibowo	0.593639	3
	35 Safrian Aswati	0.592554	8
	32 Rida Indah Fariani	0.591704	3

Model Stemming 50 Topik: Dokumen 4

	Author	Score	Size
43	Wiwik Anggraeni	0.597509	5
2	Andeka Rocky Tanaamah	0.594031	3
7	Arif Wibisono	0.593833	3
41	Wayan Firdaus Mahmudy	0.593781	6
31	Renny P. Kusumawardani	0.593451	4
22	Iskandar	0.590365	3
45	Zulfi Azhar	0.588682	3
20	Hanim Maria Astuti	0.586776	5
16	Febriliyan Samopa	0.586117	5
37	Stanley Karouw	0.585642	3

Model Stemming 50 Topik: Dokumen 5

	Author	Score	Size
7	Arif Wibisono	0.592943	3
41	Wayan Firdaus Mahmudy	0.592832	6
31	Renny P. Kusumawardani	0.592562	4
2	Andeka Rocky Tanaamah	0.592121	3
43	Wiwik Anggraeni	0.591530	5
22	Iskandar	0.589409	3
20	Hanim Maria Astuti	0.586925	5
16	Febriliyan Samopa	0.586365	5
37	Stanley Karouw	0.586088	3
27	Nisfu Asrul Sani	0.586036	3

Model Stemming 50 Topik: Dokumen 6

	Author	Score	Size
33	Royyana M. Ijtihadie	0.589423	3
24	Mira Ziveria	0.589281	3
11	Cut Fiarni	0.587134	4
45	Zulfi Azhar	0.586630	3
37	Stanley Karouw	0.586510	3
27	Nisfu Asrul Sani	0.586458	3
42	William Ramdhan	0.586437	3
12	Desy Iba Ricoida	0.586334	3
7	Arif Wibisono	0.586099	3
32	Rida Indah Fariani	0.586047	3

Model Stemming 50 Topik: Dokumen 7

	Author	Score	Size
20	Hanim Maria Astuti	0.592681	5
43	Wiwik Anggraeni	0.588041	5
45	Zulfi Azhar	0.587930	3
37	Stanley Karouw	0.586219	3
27	Nisfu Asrul Sani	0.586168	3
42	William Ramdhan	0.586147	3
12	Desy Iba Ricoida	0.586044	3
7	Arif Wibisono	0.585809	3
32	Rida Indah Fariani	0.585757	3
3	Andre Parvian Aristio	0.585719	3

Model Stemming 50 Topik: Dokumen 8

	Author	Score	Size
16	Febriliyan Samopa	0.591970	5
42	William Ramdhan	0.591759	3
29	Radityo Prasetyanto Wibowo	0.588404	3
37	Stanley Karouw	0.586076	3
27	Nisfu Asrul Sani	0.586025	3
12	Desy Iba Ricoida	0.585901	3
7	Arif Wibisono	0.585666	3
32	Rida Indah Fariani	0.585615	3
3	Andre Parvian Aristio	0.585576	3
28	Oleh Soleh	0.585473	3

Model Stemming 50 Topik: Dokumen 9

	Author	Score	Size
33	Royyana M. Ijtihadie	0.593516	3
37	Stanley Karouw	0.593315	3
28	Oleh Soleh	0.592801	3
21	Henning Titi Ciptaningtyas	0.592790	4
32	Rida Indah Fariani	0.589195	3
24	Mira Ziveria	0.586952	3
27	Nisfu Asrul Sani	0.585986	3
42	William Ramdhan	0.585965	3
12	Desy Iba Ricoida	0.585862	3
7	Arif Wibisono	0.585628	3

Model Stemming 50 Topik: Dokumen 10

	Author	Score	Size
45	Zulfi Azhar	0.590556	3
11	Cut Fiarni	0.588021	4
14	Erma Suryani	0.586754	8
10	Bekti Cahyo Hidayanto	0.586092	3
37	Stanley Karouw	0.585883	3
27	Nisfu Asrul Sani	0.585832	3
42	William Ramdhan	0.585811	3
12	Desy Iba Ricoida	0.585708	3
7	Arif Wibisono	0.585474	3
32	Rida Indah Fariani	0.585422	3

Model Stemming 50 Topik: Dokumen 11

	Author	Score	Size
43	Wiwik Anggraeni	0.596887	5
7	Arif Wibisono	0.593872	3
2	Andeka Rocky Tanaamah	0.593838	3
41	Wayan Firdaus Mahmudy	0.593811	6
31	Renny P. Kusumawardani	0.593490	4
22	Iskandar	0.590346	3
45	Zulfi Azhar	0.588563	3
20	Hanim Maria Astuti	0.586968	5
16	Febriliyan Samopa	0.586323	5
37	Stanley Karouw	0.585878	3

Model Stemming 50 Topik: Dokumen 12

	Author	Score	Size
26	Neni Mulyani	0.596860	3
8	Arridha Zikra Syah	0.596167	3
45	Zulfi Azhar	0.587321	3
27	Nisfu Asrul Sani	0.587242	3
7	Arif Wibisono	0.586894	3
34	Rully Agus Hendrawan	0.586624	5
37	Stanley Karouw	0.585956	3
42	William Ramdhan	0.585884	3
12	Desy Iba Ricoida	0.585781	3
32	Rida Indah Fariani	0.585495	3

Model Stemming 50 Topik: Dokumen 13

	Author	Score	Size
	7 Arif Wibisono	0.591016	3
41	Wayan Firdaus Mahmudy	0.590808	6
31	Renny P. Kusumawardani	0.590638	4
	6 Arif Djunaidy	0.590613	7
2	Andeka Rocky Tanaamah	0.589949	3
15	Evasaria M. Sipayung	0.589937	4
43	Wiwik Anggraeni	0.589572	5
	22 Iskandar	0.588243	3
10	Bekti Cahyo Hidayanto	0.587924	3
40	Triyanna Widyaningtyas	0.587693	3

Model Stemming 50 Topik: Dokumen 14

	Author	Score	Size
20	Hanim Maria Astuti	0.594722	5
45	Zulfi Azhar	0.588163	3
43	Wiwik Anggraeni	0.587911	5
37	Stanley Karouw	0.586133	3
27	Nisfu Asrul Sani	0.586081	3
42	William Ramdhan	0.586060	3
12	Desy Iba Ricoida	0.585957	3
7	Arif Wibisono	0.585723	3
32	Rida Indah Fariani	0.585671	3
3	Andre Parvian Aristio	0.585633	3

Model Stemming 50 Topik: Dokumen 15

	Author	Score	Size
36	Sofu Risqi Y.S.	0.593591	3
13	Ely Rosely	0.593503	3
23	Mahendrawathi ER	0.593405	3
7	Arif Wibisono	0.593338	3
41	Wayan Firdaus Mahmudy	0.593249	6
31	Renny P. Kusumawardani	0.592957	4
34	Rully Agus Hendrawan	0.592920	5
14	Erma Suryani	0.592752	8
45	Zulfi Azhar	0.588758	3
20	Hanim Maria Astuti	0.586939	5

Model Stemming 50 Topik: Dokumen 16

	Author	Score	Size
26	Neni Mulyani	0.599424	3
8	Arridha Zikra Syah	0.599326	3
30	Raras Tyasnurita	0.589707	3
17	Feby Artwodini Muqtadiroh	0.589605	8
18	Fransiskus Adikara	0.589603	4
29	Radityo Prasetianto Wibowo	0.589348	3
42	William Ramdhan	0.588247	3
27	Nisfu Asrul Sani	0.587526	3
45	Zulfi Azhar	0.587314	3
7	Arif Wibisono	0.587179	3

Model Stemming 50 Topik: Dokumen 17

	Author	Score	Size
37	Stanley Karouw	0.591802	3
36	Sofu Risqi Y.S.	0.591426	3
13	Ely Rosely	0.591338	3
28	Oleh Soleh	0.591265	3
23	Mahendrawathi ER	0.591241	3
38	Titus Kristanto	0.591120	4
20	Hanim Maria Astuti	0.590094	5
32	Rida Indah Fariani	0.588591	3
24	Mira Ziveria	0.586817	3
0	Alz Danny Wowor	0.586395	4

Model Stemming 50 Topik: Dokumen 18

	Author	Score	Size
7	Arif Wibisono	0.590401	3
41	Wayan Firdaus Mahmudy	0.590167	6
6	Arif Djunaidy	0.590148	7
31	Renny P. Kusumawardani	0.590023	4
15	Evasaria M. Sipayung	0.589541	4
10	Bekti Cahyo Hidayanto	0.587717	3
43	Wiwik Anggraeni	0.587550	5
40	Triyanna Widyaningtyas	0.587513	3
3	Andre Parvian Aristio	0.587225	3
20	Hanim Maria Astuti	0.586619	5

Model Stemming 50 Topik: Dokumen 19

	Author	Score	Size
36	Sofu Risqi Y.S.	0.592676	3
13	Ely Rosely	0.592588	3
23	Mahendrawathi ER	0.592491	3
7	Arif Wibisono	0.591325	3
41	Wayan Firdaus Mahmudy	0.591134	6
31	Renny P. Kusumawardani	0.590946	4
20	Hanim Maria Astuti	0.586758	5
0	Alz Danny Wowor	0.586600	4
37	Stanley Karouw	0.586340	3
16	Febriliyan Samopa	0.586340	5

Model Stemming 50 Topik: Dokumen 20

	Author	Score	Size
43	Wiwik Anggraeni	0.596972	5
2	Andeka Rocky Tanaamah	0.594564	3
7	Arif Wibisono	0.593224	3
41	Wayan Firdaus Mahmudy	0.593126	6
31	Renny P. Kusumawardani	0.592842	4
0	Alz Danny Wowor	0.592568	4
22	Iskandar	0.590852	3
20	Hanim Maria Astuti	0.586969	5
16	Febriliyan Samopa	0.586387	5
37	Stanley Karouw	0.586065	3

D-4. Hasil Pengujian *Stemming* 50 Topik – *Minimal Author Size=5*

Model *Stemming* 50 Topik: Dokumen 1

	Author	Score	Size
9	Rully Agus Hendrawan	0.594052	5
	4 Erma Suryani	0.593864	8
	10 Safrian Aswati	0.585816	8
	13 Wiwik Anggraeni	0.584920	5
	14 Yessica Siagian	0.584887	5
	8 Hanim Maria Astuti	0.584887	5
7	Guardian Yoki Sanjaya	0.584884	6
	0 Amna Shifia Nisafani	0.584882	6
	5 Febriliyan Samopa	0.584881	5
12	Wayan Firdaus Mahmudy	0.584874	6

Model *Stemming* 50 Topik: Dokumen 2

	Author	Score	Size
	13 Wiwik Anggraeni	0.585663	5
9	Rully Agus Hendrawan	0.585657	5
	14 Yessica Siagian	0.585630	5
	8 Hanim Maria Astuti	0.585630	5
7	Guardian Yoki Sanjaya	0.585627	6
	0 Amna Shifia Nisafani	0.585625	6
	5 Febriliyan Samopa	0.585624	5
12	Wayan Firdaus Mahmudy	0.585617	6
6	Feby Artwodini Muqtadiroh	0.585613	8
	2 Apol Pribadi Subriadi	0.585610	7

Model *Stemming* 50 Topik: Dokumen 3

	Author	Score	Size
	8 Hanim Maria Astuti	0.599362	5
6	Feby Artwodini Muqtadiroh	0.594351	8
	10 Safrian Aswati	0.592554	8
	13 Wiwik Anggraeni	0.584894	5
9	Rully Agus Hendrawan	0.584887	5
	14 Yessica Siagian	0.584861	5
7	Guardian Yoki Sanjaya	0.584858	6
	0 Amna Shifia Nisafani	0.584856	6
	5 Febriliyan Samopa	0.584855	5
12	Wayan Firdaus Mahmudy	0.584848	6

Model Stemming 50 Topik: Dokumen 4

	Author	Score	Size
13	Wiwik Anggraeni	0.597509	5
12	Wayan Firdaus Mahmudy	0.593781	6
8	Hanim Maria Astuti	0.586776	5
5	Febriliyan Samopa	0.586117	5
9	Rully Agus Hendrawan	0.584862	5
14	Yessica Siagian	0.584835	5
7	Guardian Yoki Sanjaya	0.584833	6
0	Amna Shifia Nisafani	0.584830	6
6	Feby Artwodini Muqtadiroh	0.584818	8
2	Apol Pribadi Subriadi	0.584815	7

Model Stemming 50 Topik: Dokumen 5

	Author	Score	Size
12	Wayan Firdaus Mahmudy	0.592832	6
13	Wiwik Anggraeni	0.591530	5
8	Hanim Maria Astuti	0.586925	5
5	Febriliyan Samopa	0.586365	5
9	Rully Agus Hendrawan	0.585306	5
14	Yessica Siagian	0.585279	5
7	Guardian Yoki Sanjaya	0.585276	6
0	Amna Shifia Nisafani	0.585274	6
6	Feby Artwodini Muqtadiroh	0.585262	8
2	Apol Pribadi Subriadi	0.585258	7

Model Stemming 50 Topik: Dokumen 6

	Author	Score	Size
13	Wiwik Anggraeni	0.585731	5
9	Rully Agus Hendrawan	0.585725	5
14	Yessica Siagian	0.585698	5
8	Hanim Maria Astuti	0.585698	5
7	Guardian Yoki Sanjaya	0.585695	6
0	Amna Shifia Nisafani	0.585693	6
5	Febriliyan Samopa	0.585692	5
12	Wayan Firdaus Mahmudy	0.585685	6
6	Feby Artwodini Muqtadiroh	0.585681	8
2	Apol Pribadi Subriadi	0.585678	7

Model Stemming 50 Topik: Dokumen 7

	Author	Score	Size
8	Hanim Maria Astuti	0.592681	5
13	Wiwik Anggraeni	0.588041	5
9	Rully Agus Hendrawan	0.585436	5
14	Yessica Siagian	0.585410	5
7	Guardian Yoki Sanjaya	0.585407	6
0	Amna Shifia Nisafani	0.585404	6
5	Febriliyan Samopa	0.585404	5
12	Wayan Firdaus Mahmudy	0.585397	6
6	Feby Artwodini Muqtadiroh	0.585393	8
2	Apol Pribadi Subriadi	0.585389	7

Model Stemming 50 Topik: Dokumen 8

	Author	Score	Size
5	Febriliyan Samopa	0.591970	5
13	Wiwik Anggraeni	0.585300	5
9	Rully Agus Hendrawan	0.585294	5
14	Yessica Siagian	0.585267	5
8	Hanim Maria Astuti	0.585267	5
7	Guardian Yoki Sanjaya	0.585265	6
0	Amna Shifia Nisafani	0.585262	6
12	Wayan Firdaus Mahmudy	0.585255	6
6	Feby Artwodini Muqtadiroh	0.585250	8
2	Apol Pribadi Subriadi	0.585247	7

Model Stemming 50 Topik: Dokumen 9

	Author	Score	Size
13	Wiwik Anggraeni	0.585262	5
9	Rully Agus Hendrawan	0.585255	5
14	Yessica Siagian	0.585229	5
8	Hanim Maria Astuti	0.585229	5
7	Guardian Yoki Sanjaya	0.585226	6
0	Amna Shifia Nisafani	0.585223	6
5	Febriliyan Samopa	0.585223	5
12	Wayan Firdaus Mahmudy	0.585216	6
6	Feby Artwodini Muqtadiroh	0.585212	8
2	Apol Pribadi Subriadi	0.585208	7

Model Stemming 50 Topik: Dokumen 10

	Author	Score	Size
4	Erma Suryani	0.586754	8
13	Wiwik Anggraeni	0.585108	5
9	Rully Agus Hendrawan	0.585102	5
14	Yessica Siagian	0.585075	5
8	Hanim Maria Astuti	0.585075	5
7	Guardian Yoki Sanjaya	0.585072	6
0	Amna Shifia Nisafani	0.585070	6
5	Febriliyan Samopa	0.585069	5
12	Wayan Firdaus Mahmudy	0.585063	6
6	Feby Artwodini Muqtadiroh	0.585058	8

Model Stemming 50 Topik: Dokumen 11

	Author	Score	Size
13	Wiwik Anggraeni	0.596887	5
12	Wayan Firdaus Mahmudy	0.593811	6
8	Hanim Maria Astuti	0.586968	5
5	Febriliyan Samopa	0.586323	5
9	Rully Agus Hendrawan	0.585097	5
14	Yessica Siagian	0.585071	5
7	Guardian Yoki Sanjaya	0.585068	6
0	Amna Shifia Nisafani	0.585065	6
6	Feby Artwodini Muqtadiroh	0.585054	8
2	Apol Pribadi Subriadi	0.585050	7

Model Stemming 50 Topik: Dokumen 12

	Author	Score	Size
9	Rully Agus Hendrawan	0.586624	5
13	Wiwik Anggraeni	0.585181	5
14	Yessica Siagian	0.585148	5
8	Hanim Maria Astuti	0.585148	5
7	Guardian Yoki Sanjaya	0.585145	6
0	Amna Shifia Nisafani	0.585143	6
5	Febriliyan Samopa	0.585142	5
12	Wayan Firdaus Mahmudy	0.585135	6
6	Feby Artwodini Muqtadiroh	0.585131	8
2	Apol Pribadi Subriadi	0.585128	7

Model Stemming 50 Topik: Dokumen 13

	Author	Score	Size
12	Wayan Firdaus Mahmudy	0.590808	6
	3 Arif Djunaidy	0.590613	7
13	Wiwik Anggraeni	0.589572	5
8	Hanim Maria Astuti	0.586754	5
5	Febriliyan Samopa	0.586365	5
9	Rully Agus Hendrawan	0.585640	5
	14 Yessica Siagian	0.585613	5
7	Guardian Yoki Sanjaya	0.585611	6
0	Amna Shifia Nisafani	0.585608	6
6	Feby Artwodini Muqtadiroh	0.585596	8

Model Stemming 50 Topik: Dokumen 14

	Author	Score	Size
8	Hanim Maria Astuti	0.594722	5
13	Wiwik Anggraeni	0.587911	5
9	Rully Agus Hendrawan	0.585350	5
	14 Yessica Siagian	0.585324	5
7	Guardian Yoki Sanjaya	0.585321	6
0	Amna Shifia Nisafani	0.585318	6
5	Febriliyan Samopa	0.585318	5
12	Wayan Firdaus Mahmudy	0.585311	6
6	Feby Artwodini Muqtadiroh	0.585307	8
	2 Apol Pribadi Subriadi	0.585303	7

Model Stemming 50 Topik: Dokumen 15

	Author	Score	Size
12	Wayan Firdaus Mahmudy	0.593249	6
9	Rully Agus Hendrawan	0.592920	5
	4 Erma Suryani	0.592752	8
8	Hanim Maria Astuti	0.586939	5
5	Febriliyan Samopa	0.586343	5
	10 Safrian Aswati	0.586140	8
13	Wiwik Anggraeni	0.585217	5
	14 Yessica Siagian	0.585184	5
7	Guardian Yoki Sanjaya	0.585182	6
0	Amna Shifia Nisafani	0.585179	6

Model Stemming 50 Topik: Dokumen 16

	Author	Score	Size
6	Feby Artwodini Muqtadiroh	0.589605	8
9	Rully Agus Hendrawan	0.586917	5
10	Safrian Aswati	0.586240	8
13	Wiwik Anggraeni	0.585348	5
14	Yessica Siagian	0.585316	5
8	Hanim Maria Astuti	0.585315	5
7	Guardian Yoki Sanjaya	0.585313	6
0	Amna Shifia Nisafani	0.585310	6
5	Febriliyan Samopa	0.585310	5
12	Wayan Firdaus Mahmudy	0.585303	6

Model Stemming 50 Topik: Dokumen 17

	Author	Score	Size
8	Hanim Maria Astuti	0.590094	5
10	Safrian Aswati	0.586170	8
13	Wiwik Anggraeni	0.585539	5
9	Rully Agus Hendrawan	0.585533	5
14	Yessica Siagian	0.585506	5
7	Guardian Yoki Sanjaya	0.585503	6
0	Amna Shifia Nisafani	0.585500	6
5	Febriliyan Samopa	0.585500	5
12	Wayan Firdaus Mahmudy	0.585493	6
6	Feby Artwodini Muqtadiroh	0.585489	8

Model Stemming 50 Topik: Dokumen 18

	Author	Score	Size
12	Wayan Firdaus Mahmudy	0.590167	6
3	Arif Djunaidy	0.590148	7
13	Wiwik Anggraeni	0.587550	5
8	Hanim Maria Astuti	0.586619	5
5	Febriliyan Samopa	0.586278	5
9	Rully Agus Hendrawan	0.585646	5
14	Yessica Siagian	0.585619	5
7	Guardian Yoki Sanjaya	0.585616	6
0	Amna Shifia Nisafani	0.585613	6
6	Feby Artwodini Muqtadiroh	0.585602	8

Model Stemming 50 Topik: Dokumen 19

	Author	Score	Size
12	Wayan Firdaus Mahmudy	0.591134	6
8	Hanim Maria Astuti	0.586758	5
5	Febriliyan Samopa	0.586340	5
10	Safrian Aswati	0.586338	8
13	Wiwik Anggraeni	0.585563	5
9	Rully Agus Hendrawan	0.585556	5
14	Yessica Siagian	0.585530	5
7	Guardian Yoki Sanjaya	0.585527	6
0	Amna Shifia Nisafani	0.585524	6
6	Feby Artwodini Muqtadiroh	0.585513	8

Model Stemming 50 Topik: Dokumen 20

	Author	Score	Size
13	Wiwik Anggraeni	0.596972	5
12	Wayan Firdaus Mahmudy	0.593126	6
8	Hanim Maria Astuti	0.586969	5
5	Febriliyan Samopa	0.586387	5
9	Rully Agus Hendrawan	0.585283	5
14	Yessica Siagian	0.585256	5
7	Guardian Yoki Sanjaya	0.585253	6
0	Amna Shifia Nisafani	0.585251	6
6	Feby Artwodini Muqtadiroh	0.585239	8
2	Apol Pribadi Subriadi	0.585236	7