



**ITS**  
Institut  
Teknologi  
Sepuluh Nopember

**TUGAS AKHIR - IS184853**

**VISUALISASI PENGELOMPOKKAN TOPIK STATUS  
TWITTER CALEG BERDASARKAN KUBU PENDUKUNG  
CAPRES DAN CAWAPRES DALAM PEMILU 2019  
MENGGUNAKAN AUTHOR-TOPIC MODEL**

**VISUALIZATION OF GROUPING LEGISLATIVE  
CANDIDATES TWITTER STATUS BASED ON  
PRESIDENT AND VICE PRESIDENT SUPPORTING SIDE  
IN INDONESIA 2019 ELECTION USING AUTHOR-  
TOPIC MODEL**

UMI SHOIMA  
0521154000047

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

DEPARTEMEN SISTEM INFORMASI  
Fakultas Teknologi Informasi dan Komunikasi  
Institut Teknologi Sepuluh Nopember  
Surabaya 2019



**TUGAS AKHIR - IS184853**

**VISUALISASI PENGELOMPOKAN TOPIK  
STATUS TWITTER CALEG BERDASARKAN  
KUBU PENDUKUNG CAPRES DAN CAWAPRES  
DALAM PEMILU 2019 MENGGUNAKAN  
AUTHOR-TOPIC MODEL**

**UMI SHOIMA  
0521154000047**

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

**DEPARTEMEN SISTEM INFORMASI  
Fakultas Teknologi Informasi dan Komunikasi  
Institut Teknologi Sepuluh Nopember  
Surabaya 2019**



**UNDERGRADUATE THESIS - IS184853**

**VISUALIZATION OF GROUPING LEGISLATIVE  
CANDIDATES TWITTER STATUS BASED ON  
PRESIDENT AND VICE PRESIDENT SUPPORTING  
SIDE IN INDONESIA 2019 ELECTION USING  
AUTHOR-TOPIC MODEL**

**UMI SHOIMA**  
**0521154000047**

**Supervisor**

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

**INFORMATION SYSTEM DEPARTMENT**

**Information Technology and Communication Faculty  
Sepuluh Nopember Institute of Technology  
Surabaya 2019**



# LEMBAR PENGESAHAN

## VISUALISASI PENGELOMPOKAN TOPIK STATUS TWITTER CALEG BERDASARKAN KUBU PENDUKUNG CAPRES DAN CAWAPRES DALAM PEMILU 2019 MENGUNAKAN AUTHOR-TOPIC

### 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:

**UMI SHOIMA**  
**0521154000047**

Surabaya, Juli 2019

**KEPALA**  
**DEPARTEMEN SISTEM INFORMASI**



  
**Mahendrawati ER, ST, M.Sc, P.hd**  
**NIP 197610112 006042 001**





## LEMBAR PERSETUJUAN

# ***VISUALISASI PENGELOMPOKKAN TOPIK STATUS TWITTER CALEG BERDASARKAN KUBU PENDUKUNG CAPRES DAN CAWAPRES DALAM PEMILU 2019 MENGGUNAKAN AUTHOR-TOPIC***

### 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:

**UMI SHOIMA**  
**0521154000047**

Disetujui Tim Penguji: Tanggal Ujian : 11 Juli 2019  
Periode Wisuda : September 2019

**Nur Aini R., S.Kom., M.Sc. Eng., Ph.D**

(Pembimbing 1)

**Faisal Johan Atletiko, S.Kom, M.Kom**

(Penguji 1)

**Irmasari Hafidz, S.Kom, M.Sc**

(Penguji 2)





**VISUALISASI PENGELOMPOKAN TOPIK STATUS  
TWITTER CALEG BERDASARKAN KUBU  
PENDUKUNG CAPRES DAN CAWAPRES DALAM  
PEMILU 2019 MENGGUNAKAN AUTHOR-TOPIC  
MODEL**

**Nama Mahasiswa** : Umi Shoima  
**NRP** : 0521154000047  
**Departemen** : Sistem Informasi FTIK-ITS  
**Pembimbing I** : Nur Aini Rakhmawati, S.Kom., M.Sc.  
Eng., Ph.D

**ABSTRAK**

*Pada era teknologi yang semakin berkembang, menggunakan media sosial sebagai kampanye adalah hal yang banyak dipilih oleh calon legislatif. Salah satu media sosial yang banyak dipilih oleh para calon legislatif untuk menyampaikan pendapat, dukungan, maupun kritikan demi mendapatkan perhatian dan dukungan dari masyarakat adalah Twitter. Namun dari setiap tweet yang dibagikan oleh para calon legislatif, masih belum dapat dipetakan dan dikelompokkan topik apa saja yang menjadi bahasan atau fokus dari masing-masing kubu pendukung pasangan calon presiden dan calon wakil presiden. Untuk itu, diperlukan sebuah penelitian untuk memodelkan topik dari tweet para calon legislatif, dengan metode yang dipilih adalah menggunakan Author-Topic Model. Metode ini dapat mengklasifikasikan tweet yang sudah didapatkan dan diolah menjadi beberapa topik beserta siapa author yang mungkin mengemukakan topik tersebut. Dengan adanya penelitian ini, diharapkan dapat membantu memetakan dan memvisualisasikan topik bahasan dari masing-masing tweet yang dibagikan oleh setiap akun calon legislatif, yang kemudian dapat merepresentasikan topik bahasan dari masing-masing kubu pasangan calon presiden dan calon wakil presiden. Dari permodelan topik yang dilakukan menggunakan Author-Topic Model, didapatkan 4 topik berdasarkan analisis nilai perplexity yaitu topik 'Isu Hukum', 'Pergerakan*

*Nasional*, *'Kebijakan Pemerintah'*, dan *'Isu Keamanan'*. Topik *'Isu Hukum'* dibahas oleh 6 partai, *'Isu Keamanan'* dibahas oleh 2 partai, topik *'Kebijakan Pemerintah'* dibahas oleh 2 partai, dan topik *'Pergerakan Nasional'* dibahas oleh 5 partai. Kemudian pada permodelan topik yang dilakukan dengan LDA atau yang digunakan untuk mengetahui topik yang dibahas pada kedua kubu capres dan cawapres, didapatkan model dengan 3 topik pada masing-masing kubu. Model tersebut juga didapatkan dengan menganalisis hasil perplexity yang didapatkan pada setiap percobaan pembuatan model topik. Caleg pada kubu 01 diketahui membahas topik yang berkaitan dengan *'Politik dan Keagamaan'*, *'Kerakyatan'*, dan *'Hoax'*. Sementara caleg pada kubu 02, membahas topik berkaitan dengan *'Kemasyarakatan'*, *'Kampanye Politik'*, dan *'Isu Kenegaraan'*.

***Kata Kunci: Author-Topic Model, LDA, Twitter, Calon Legislatif, Partai Politik, Pemilu 2019.***

# VISUALIZATION OF GROUPING LEGISLATIVE CANDIDATES TWITTER STATUS BASED ON PRESIDENT AND VICE PRESIDENT SUPPORTING SIDE IN INDONESIA 2019 ELECTION USING AUTHOR-TOPIC MODEL

**Name** : Umi Shoima  
**NRP** : 0521154000047  
**Department** : Information Systems FTIK-ITS  
**Supervisor** : Nur Aini Rakhmawati, S.Kom., M.Sc. Eng., Ph.D

## ABSTRACT

*In an increasingly evolving technological era, using social media as a campaign is something that many legislative candidates choose. One of the many social media chosen by the legislative candidates to express their opinions, support, and criticism in order to get attention and support from the public is Twitter. However, from every tweet shared by the legislative candidates, it still cannot be mapped and grouped any topics that are the subject of discussion or the focus from each of the supporters of the presidential and vice-presidential candidate supporters. For this reason, we need a study to model the topic of the tweets of legislative candidates, with the chosen method to use is Author-Topic Model. This method can classify tweets that have been obtained and processed into several topics along with who the author might present the topic. With this research, it is hoped that it can help map and visualize the discussion topics of each tweet shared by each candidate's legislative account, which can then represent the topic of discussion of each of the presidential and vice-presidential candidate pairs. From topic modeling carried out using the Author-Topic Model, there were 4 topics based on analysis of perplexity values, namely the topic of 'Legal Issues', 'National Movement', 'Government Policy', and 'Security Issues'. The topic of the 'Legal Issues' was discussed by 6 parties, 'Security Issues' discussed by 2 parties, the topic 'Government Policy' discussed*

*by 2 parties, and the topic 'National Movement' was discussed by 5 parties. Then on the topic modeling carried out with LDA or used to find out the topics discussed in both the presidential and vice presidential camps, a model with 3 topics was obtained from each side. The model was also obtained by analyzing the results of perplexity obtained in each topic modeling experiment. Candidate from 01 side is known to discuss topics related to 'Politics and Religion', 'Population', and 'Hoax'. While the candidates from 02 side discussed topics related to 'Community', 'Political Campaign', and 'State Issues'.*

***Keywords: Author-Topic Model, LDA, Twitter, Legislative Candidates, Political Parties, 2019 Election.***

## KATA PENGANTAR

Dengan mengucapkan rasa syukur kepada Allah SWT Yang Maha Pengasih dan Maha Penyayang atas izin-Nya penulis dapat menyelesaikan buku tugas akhir ini dengan judul “Visualisasi Pengelompokan Topik Status Twitter Caleg berdasarkan Kubu Pendukung Capres dan Cawapres dalam Pemilu 2019 menggunakan Author-Topic Model”. Dalam penyelesaian Tugas Akhir ini, penulis diiringi oleh pihak-pihak yang selalu memberi dukungan, saran, dan doa sehingga penelitian berlangsung dengan lancar. Secara khusus penulis mengucapkan terima kasih kepada:

1. Allah SWT, yang telah memberikan kekuatan, hidayah, serta ilmu yang memungkinkan penulis bisa menyelesaikan tugas akhir ini dengan semaksimal mungkin.
2. Bapak Sri Adi, Almh. Ibu Munawaroh, dan Adik Muzaki yang telah memberikan dukungan, semangat, doa, dan motivasi bagi penulis untuk bisa menyelesaikan studi S1 dengan sebaik-baiknya.
3. Ibu Mahendrawati ER, ST, M.Sc, Ph.D selaku Kadep Sistem Informasi ITS, Bapak Nisfu Asrul Sani, S.Kom, M.Sc selaku Kaprodi S1 Sistem Informasi ITS, serta seluruh dosen pengajar beserta staf dan karyawan di Departemen Sistem Informasi, FTIK ITS Surabaya selama penulis menjalani kuliah.
4. Ibu Nur Aini Rakhmawati, S.Kom., M.Sc., Eng. Ph.D selaku dosen pembimbing selama pengerjaan tugas akhir ini yang telah meluangkan waktu dan pikiran demi membantu terselesaikannya tugas akhir ini dengan baik.
5. Bapak Faizal Johan Atletiko, S.Kom., M.T dan Ibu Irmasari Hafidz, S.Kom, M.Sc selaku dosen penguji yang telah memberikan saran, masukan yang membuat tugas akhir ini bisa lebih baik lagi.
6. Teman-teman saya sejak kelas XI SMA yaitu Tataques Friends yang terdiri dari Bag, Henqi, dan Mama yang telah

menemani, mendukung, dan memaksa penulis agar segera menyelesaikan tugas akhir ini.

7. Anggota grup Waspada 4 M yang kadang di Line, kadang di Whatsapp, yaitu Cincinia, Palupi, dan Dinan yang telah menemani penulis selama masa perkuliahan.
8. Teman-teman grup sudah lengkap yang kadang-kadang saling miskomunikasi tapi pada akhirnya semua baik-baik saja.
9. Keluarga besar Lannister, yang telah menemani penulis sejak masa maba hingga sekarang.
10. Adik KP Mulmulidelicious yang telah menemani hari-hari penulis.
11. Seluruh teman-teman bimbingan Bu Iin yang telah bersama-sama berjuang sejak masa awal pendaftaran bimbingan hingga masa sidang akhir.
12. Teman-teman ADDI yang telah bersama-sama berjuang menyelesaikan tugas akhir hingga selesai.

Penyusunan tugas akhir ini masih jauh dari kata sempurna, untuk itu penulis menerima segala kritik dan saran yang membangun sebagai upaya menjadi lebih baik lagi ke depannya. Semoga tugas akhir ini dapat memberikan manfaat kedepannya.

Surabaya, Juli 2019

Penulis



## DAFTAR ISI

ABSTRAK .....	i
ABSTRACT .....	iii
KATA PENGANTAR .....	v
DAFTAR ISI .....	vii
DAFTAR GAMBAR .....	xi
DAFTAR TABEL .....	xiii
DAFTAR KODE .....	xv
DAFTAR GRAFIK .....	xvii
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	3
1.3 Batasan Permasalahan .....	4
1.4 Tujuan .....	4
1.5 Manfaat .....	4
1.6 Relevansi .....	5
BAB II TINJAUAN PUSTAKA .....	7
2.1 Penelitian Terkait .....	7
2.2 Dasar Teori .....	9
2.2.1 Twitter .....	9
2.2.2 Pemilihan Umum 2019 .....	9
2.2.3 Social Media Crawler .....	12
2.2.4 Topic Model .....	13
2.2.5 Latent Dirichlet Allocation .....	13
2.2.6 Author-Topic Model .....	14
2.2.7 Perplexity .....	17
2.2.8 Tableau .....	17
BAB III METODOLOGI .....	19
3.1 Tahapan Pelaksanaan Tugas Akhir .....	19
3.1.1 Studi Literatur .....	20
3.1.2 Akuisisi Data .....	20
3.1.3 Data Pre-processing .....	22
3.1.4 Permodelan Topik .....	25
3.1.5 Pembuatan Visualisasi Topik Pilpres 2019 ....	25
BAB IV PERANCANGAN .....	27

4.1	Arsitektur Sistem.....	27
4.2	Pengambilan Data .....	28
4.3	Desain Basis Data Penyimpanan Tweet dan Akun Twitter 29	
4.4	Data Pre-Processing .....	30
4.5	Pemrosesan Data .....	31
	4.5.1 Pembentukan Author-to-doc .....	31
	4.5.2 Pembentukan Dictionary dan Corpus.....	32
	4.5.3 Pembentukan Topic Model .....	32
4.6	Validasi Permodelan Topik.....	33
4.7	Analisis Topik.....	34
4.8	Desain Visualisasi Permodelan Topik .....	34
<b>BAB V IMPLEMENTASI .....</b>		<b>37</b>
5.1	Lingkungan Implementasi.....	37
5.2	Pengambilan dan Penyimpanan Data pada Basis Data 38	
	5.2.1 Pengumpulan Data Akun .....	38
	5.2.2 Crawling Data .....	40
5.3	Data Pre-Processing .....	42
5.4	Pemrosesan Data .....	47
	5.4.1 Proses Memuat Data .....	47
	5.4.2 Proses Pembuatan Dictionary dan Corpus .....	48
	5.4.3 Pencatatan Nilai Perplexity .....	49
	5.4.4 Pembentukan Topik dengan Author Kubu Pendukung Capres dan Cawapres .....	50
	5.4.5 Pembentukan Topik dengan Author Partai .....	51
	5.4.6 Penyimpanan Model Topik .....	53
5.5	Validasi Permodelan Topik.....	54
5.6	Analisis Topik.....	54
5.7	Visualisasi Permodelan Topik.....	56
<b>BAB VI HASIL DAN PEMBAHASAN.....</b>		<b>65</b>
6.1	Hasil Tweet secara Umum .....	65
6.2	Pre-processing Data .....	68
6.3	Permodelan Topik dan Validasi.....	69
	6.3.1 Permodelan Topik dengan Author-Topic Model 69	
	6.3.2 Permodelan Topik dengan Latent Dirichlet Allocation.....	86
<b>BAB VII KESIMPULAN DAN SARAN .....</b>		<b>107</b>

7.1 Kesimpulan .....	107
7.2 Saran .....	108
DAFTAR PUSTAKA .....	109
LAMPIRAN A. Hasil Pengujian Model menggunakan ATM dan LDA.....	113
LAMPIRAN B. Link Profil Tableau Public Visualisasi Permodelan Topik.....	113
LAMPIRAN C. List Stopwords.....	113
BIODATA PENULIS .....	115

*Halaman sengaja dikosongkan.*

## DAFTAR GAMBAR

Gambar 2.1 Arsitektur Web Crawler .....	12
Gambar 2.2 Konsep Permodelan Topik[19].....	13
Gambar 2.3 Konsep Latent Dirichlet Allocation[20].....	14
Gambar 2.4 Konsep Author Topic Model[21] .....	15
Gambar 2.5 Contoh Topik dari Model Author-Topic dengan dataset Neural Information Processing Systems (NIPS) Conference berdasarkan topik[21] .....	16
Gambar 2.6 Contoh Topik dari Model Author-Topic dengan dataset Neural Information Processing Systems (NIPS) Conference berdasarkan author[21] .....	16
Gambar 2.7 Contoh Tampilan Tableau Desktop.....	17
Gambar 2.8 Contoh Tampilan Tableau Public.....	18
Gambar 3.1 Metodologi Penelitian .....	19
Gambar 3.2 Proses Validasi Akun Twitter Caleg .....	21
Gambar 3.3 Proses Pengumpulan data status Twitter .....	21
Gambar 3.4 Pengelompokkan Caleg untuk Permodelan Topik .....	26
Gambar 4.1 Arsitektur Sistem.....	27
Gambar 4.2 Desain Penyimpanan Basis Data .....	29
Gambar 4.3 Pre-processing Tweet .....	30
Gambar 4.4 Gambaran Umum Pembuatan Model Topik.....	31
Gambar 4.5 Proses Pembuatan Author-to-doc.....	32
Gambar 4.6 Proses Pembentukan Dictionary dan Corpus .....	32
Gambar 4.7 Proses Permodelan Topik dengan ATM dan LDA .....	33
Gambar 4.8 Desain Visualisasi Author pada tiap Topik .....	35
Gambar 4.9 Desain Visualisasi Detail Kata tiap Topik.....	35
Gambar 5.1 Detail Sheet Statistik Tweet .....	57
Gambar 5.2 Detail Sheet Total Tweet per Partai.....	57
Gambar 5.3 Detail Sheet Perbandingan Jumlah Tweet per Partai .....	58
Gambar 5.4 Dashboard Analisis Tweet secara Umum.....	58
Gambar 5.5 Detail Sheet Hasil Topik per Partai .....	59
Gambar 5.6 Detail Sheet Detail Kata per Topik .....	60
Gambar 5.7 Dashboard Hasil Permodelan Topik tiap Partai .	61

Gambar 5.8 Detail Sheet Topik per Kubu .....	61
Gambar 5.9 Detail Sheet Detail Topik tiap Kubu .....	62
Gambar 5.10 Dashboard Hasil Permodelan Topik tiap Kubu	63
Gambar 5.11 Detail Workbook Word Cloud .....	63
Gambar 5.12 Hasil Word Cloud.....	64
Gambar 6.1 Statistik Tweet .....	66
Gambar 6.2 Perbandingan Total Tweet pada tiap Kubu .....	67
Gambar 6.3 Perbandingan Tweet pada tiap Partai .....	68
Gambar 6.4 Radio Button untuk memilih Kategori Topik.....	78
Gambar 6.5 Detail Kata pada Topik Isu Hukum .....	78
Gambar 6.6 Detail Kata pada Topik Isu Keamanan .....	80
Gambar 6.7 Detail Kata pada Topik Kebijakan Pemerintah ..	81
Gambar 6.8 Detail Kata pada Topik Pergerakan Nasional.....	83
Gambar 6.9 Detail Author Partai pada Topik Isu Hukum.....	84
Gambar 6.10 Detail Author Partai pada Topik Isu Keamanan	85
Gambar 6.11 Detail Author Partai pada Topik Kebijakan Pemerintah.....	85
Gambar 6.12 Detail Author Partai pada Topik Pergerakan Nasional.....	86
Gambar 6.13 Hasil Topik berdasarkan Kubu Capres dan Cawapres (LDA) .....	98
Gambar 6.14 Detail Kata pada Topik Hoax .....	99
Gambar 6.15 Detail Kata pada Topik Politik dan Keagamaan .....	100
Gambar 6.16 Detail Kata pada Topik Kerakyatan .....	101
Gambar 6.17 Detail Kata pada Topik Isu Kenegaraan.....	102
Gambar 6.18 Detail Kata pada Topik Kemasyarakatan .....	103
Gambar 6.19 Detail Kata pada Topik Kampanye Politik.....	105

## DAFTAR TABEL

Tabel 2.1 Literatur 1 .....	7
Tabel 2.2 Literatur 2 .....	8
Tabel 2.3 Literatur 3 .....	8
Tabel 2.4 Pembagian Partai berdasarkan Kubu Pendukung [14] .....	11
Tabel 3.1 Contoh Case Folding .....	22
Tabel 3.2 Contoh Stemming .....	23
Tabel 3.3 Contoh Tokenization .....	24
Tabel 3.4 Contoh Stopwords Removal .....	24
Tabel 5.1 Spesifikasi Perangkat Keras .....	37
Tabel 5.2 Spesifikasi Perangkat Lunak dan Library .....	37
Tabel 5.3 Kode Kubu Capres-Cawapres .....	38
Tabel 5.4 Detail Kode Partai .....	39
Tabel 5.5 Rekapitulasi Validasi Akun Twitter .....	40
Tabel 6.1 Data Total Tweet per Bulan .....	65
Tabel 6.2 Hasil Tweet per Partai .....	67
Tabel 6.3 Hasil Pre-processing .....	68
Tabel 6.4 Perbandingan Perplexity pada tiap Model Topik (ATM) .....	72
Tabel 6.5 Hasil Topik 0 dan Topik 1 pada Model dengan 3 Topik (ATM) .....	72
Tabel 6.6 Hasil Topik 2 pada Model dengan 3 Topik (ATM)	73
Tabel 6.7 Hasil Topik 0 dan Topik 1 pada Model dengan 4 Topik (ATM) .....	74
Tabel 6.8 Hasil Topik 2 dan Topik 3 pada Model dengan 4 Topik (ATM) .....	75
Tabel 6.9 Kategori Topik (ATM) .....	77
Tabel 6.10 Contoh Tweet pada Topik Isu Hukum .....	79
Tabel 6.11 Contoh Tweet pada Topik Isu Keamanan .....	81
Tabel 6.12 Contoh Tweet pada Topik Kebijakan Pemerintah	82
Tabel 6.13 Contoh Tweet pada Topik Pergerakan Nasional ..	83
Tabel 6.14 Perbandingan Perplexity pada tiap Model Topik (LDA-Kubu 01) .....	89
Tabel 6.15 Perbandingan Perplexity pada tiap Model Topik (LDA-Kubu 02) .....	90

Tabel 6.16 Hasil Topik 0 dan Topik 1 pada Model dengan 2 Topik (LDA-Kubu 01) .....	92
Tabel 6.17 Hasil Topik 0 dan Topik 1 pada Model dengan 3 Topik (LDA-Kubu 01) .....	92
Tabel 6.18 Hasil Topik 2 pada Model dengan 3 Topik (LDA-Kubu 01).....	93
Tabel 6.19 Hasil Topik 0 dan Topik 1 pada Model dengan 2 Topik (LDA-Kubu 02) .....	94
Tabel 6.20 Hasil Topik 0 dan Topik 1 pada Model dengan 3 Topik (LDA-Kubu 02) .....	95
Tabel 6.21 Hasil Topik 2 pada Model dengan 3 Topik (LDA-Kubu 02).....	96
Tabel 6.22 Kategori Topik (LDA).....	97
Tabel 6.23 Contoh Tweet pada Topik Hoax .....	99
Tabel 6.24 Contoh Tweet pada Topik Politik dan Keagamaan .....	100
Tabel 6.25 Contoh Tweet pada Topik Kerakyatan.....	101
Tabel 6.26 Contoh Tweet pada Topik Isu Kenegaraan .....	103
Tabel 6.27 Contoh Tweet pada Topik Kemasyarakatan.....	104
Tabel 6.28 Contoh Tweet pada Topik Kampanye Politik ....	105



## DAFTAR KODE

Kode 5.1 Kode untuk melakukan Crawling Tweet .....	41
Kode 5.2 Fungsi untuk melakukan Crawling .....	42
Kode 5.3 Kode untuk Lowercase .....	42
Kode 5.4 Kode untuk Penghapusan Karakter, Hashtag, Link, dan Emoji .....	43
Kode 5.5 Kode untuk Stemming .....	43
Kode 5.6 Kode untuk Tokenization .....	44
Kode 5.7 Kode untuk menghapus Stopword.....	44
Kode 5.8 Pre-processing secara keseluruhan .....	46
Kode 5.9 Fungsi untuk melakukan Pre-processing .....	47
Kode 5.10 Kode untuk memuat tweet .....	48
Kode 5.11 Kode untuk membuat Dictionary.....	49
Kode 5.12 Kode Pembentukan Corpus .....	49
Kode 5.13 Kode Pencatatan <i>Perplexity</i> .....	50
Kode 5.14 Kode Permodelan Topik dengan LDA .....	51
Kode 5.15 Kode Pembuatan Author-to-doc .....	52
Kode 5.16 Kode Permodelan Topik dengan ATM.....	53
Kode 5.17 Kode Penyimpanan Model Topik LDA.....	53
Kode 5.18 Kode Penyimpanan Model Topik ATM.....	53
Kode 5.19 Kode untuk mengambil Nilai <i>Perplexity</i> .....	54
Kode 5.20 Kode Penyimpanan Detail Kata-Kata pada Topik .....	55
Kode 5.21 Kode Penyimpanan Author pada Topik .....	56

*Halaman sengaja dikosongkan.*

## DAFTAR GRAFIK

Grafik 6.1 Perbandingan Perplexity pada tiap Passes (ATM)	70
Grafik 6.2 Rata-Rata Perplexity ATM .....	71
Grafik 6.3 Standar Deviasi Perplexity ATM.....	71
Grafik 6.4 Perbandingan Perplexity pada tiap Passes (LDA-Kubu 01).....	87
Grafik 6.5 Perbandingan Perplexity pada tiap Passes (LDA-Kubu 02).....	87
Grafik 6.6 Rata-Rata Perplexity LDA-Kubu 01.....	88
Grafik 6.7 Rata-Rata Perplexity LDA-Kubu 02.....	88
Grafik 6.8 Standar Deviasi Perplexity LDA-Kubu 01 .....	89
Grafik 6.9 Standar Deviasi Perplexity LDA-Kubu 02 .....	89
Grafik 6.10 Pengaruh Jumlah Tweet pada Nilai Perplexity ...	91

*Halaman sengaja dikosongkan.*

# **BAB I**

## **PENDAHULUAN**

Bab ini akan menjelaskan tentang pendahuluan pengerjaan tugas akhir yang meliputi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, dan manfaat yang akan diperoleh dari penelitian tugas akhir ini.

### **1.1 Latar Belakang**

Tahun 2019 akan menjadi tahun yang berpengaruh pada kehidupan berdemokrasi masyarakat Indonesia. Berdasarkan jadwal yang dibagikan oleh Komisi Pemilihan Umum, pada tanggal 17 April 2019 akan diadakan pemilihan umum, dengan pemilihan calon legislatif dan calon presiden dilakukan secara bersamaan. Salah satu kegiatan yang tidak dapat dipisahkan dari pemilihan umum adalah kampanye, yang dimulai sejak 23 September 2018 hingga 13 April 2019[1]. Menurut Undang-Undang no. 7 tahun 2017 tentang Pemilihan Umum, kampanye adalah kegiatan peserta pemilu atau pihak lain yang ditunjuk oleh peserta pemilu untuk meyakinkan pemilih dengan menawarkan visi, misi, program dan/atau citra diri peserta pemilu[2]. Dengan melakukan kampanye, peserta pemilu dapat mengemukakan berbagai rencana maupun program yang akan dilaksanakan jika mereka terpilih dalam Pemilu 2019.

Dalam melakukan kampanye, banyak cara yang bisa dilakukan oleh peserta pemilu, salah satunya adalah dengan menggunakan media sosial. Media sosial menjadi daya tarik terbaru dari para ahli strategi komunikasi politik, khususnya selama masa kampanye. Sementara televisi dan surat kabar tetap menjadi sumber berita dominan pada umumnya, media sosial telah menjadi sumber yang semakin penting bagi berita politik pada khususnya[3]. Selain itu, penggunaan media sosial sebagai media kampanye adalah untuk mengembangkan pengetahuan masyarakat mengenai isu-isu politik, yang kemudian dapat memfasilitasi partisipasi dalam kehidupan politik dan bernegara[4].

Media sosial sendiri dapat didefinisikan sebagai kumpulan aplikasi berbasis internet yang dibangun berdasarkan ideologi dan menggunakan teknologi web 2.0. Dengan aplikasi tersebut, masyarakat dapat membuat, membagikan, dan bertukar informasi melalui komunitas virtual[5]. Menurut survey yang dilakukan oleh We Are Social, total pengguna media sosial di Indonesia adalah sebanyak 130 juta pengguna, dengan rata-rata waktu yang dihabiskan untuk mengakses media sosial adalah hingga 3 jam 23 menit. Dari beberapa media sosial yang ada, salah satu media sosial yang banyak digunakan oleh masyarakat Indonesia adalah Twitter, yang menempati peringkat ketiga pada platform media sosial yang banyak diakses oleh masyarakat Indonesia[6]. Dari data tersebut, Twitter menjadi media sosial yang dipilih oleh peserta pemilu untuk melakukan kampanye. Salah satu contoh penggunaan Twitter sebagai media kampanye adalah pada saat pemilihan presiden Amerika Serikat pada tahun 2016. Dengan banyaknya pengikut dari para kandidat calon presiden, Twitter menjadi *platform* komunikasi masa dan kanal utama penyebaran informasi dari para kandidat[3].

Hal tersebut juga diikuti oleh peserta Pemilu 2019 di Indonesia. Selama masa kampanye berlangsung, para peserta pemilu 2019 yang terdiri dari calon legislatif dan pasangan capres & cawapres beserta pendukung akan membagikan *tweet* yang membahas berbagai macam hal untuk mendapatkan simpati dan dukungan masyarakat. Para calon legislatif biasanya diusung oleh sebuah partai politik, yang tergabung pada suatu koalisi tertentu. Koalisi dibentuk dengan 2 atau lebih partai politik yang mencapai tujuan masing-masing berdasarkan isu yang diprioritaskan. Selama menjadi bagian dalam suatu koalisi, setiap partai yang tergabung akan menghadapi 2 tantangan, yaitu mereka perlu menunjukkan persatuan dalam satu koalisi, tetapi pada saat yang sama, mereka harus mempertahankan dan menekankan profil mereka sendiri. Penekanan profil ini dilakukan untuk mendapatkan dukungan yang lebih banyak, dan membedakan antara satu partai dengan partai yang lain[7].

Dengan adanya perbedaan tujuan ataupun kepentingan yang diutamakan oleh masing-masing partai tetapi masih dalam ruang lingkup tujuan utama koalisi, maka tidak menutup kemungkinan adanya perbedaan topik yang dibahas selama masa kampanye[7]. Perbedaan topik yang dibahas di setiap *tweet* yang dibagikan oleh para caleg dari setiap partai yang mendukung masing-masing kubu capres & cawapres akan menggambarkan fokus utama dari tiap partai, yang lebih luasnya akan menggambarkan fokus dari tiap kubu capres & cawapres. Namun, hingga saat ini, belum ada suatu analisis yang mampu memberikan visualisasi mengenai perbedaan topik yang dibahas oleh caleg pada setiap partai yang tergabung dalam suatu koalisi. Dengan adanya visualisasi, akan membantu masyarakat mengetahui topik dan siapa caleg atau partai politik yang paling banyak membahas topik tersebut.

Sebelum dilakukan visualisasi, tentunya diperlukan analisis untuk memodelkan setiap *tweet* yang dibagikan oleh para caleg. Proses analisis akan dilakukan dengan menggunakan metode Author-Topic Model, yang merupakan pengembangan dari metode Latent Dirichlet Allocation (LDA). Metode ini dapat mengklasifikasikan *tweet* yang sudah didapatkan dan diolah menjadi beberapa topik beserta siapa *author* yang mungkin mengemukakan topik tersebut. Dengan adanya penelitian ini, diharapkan dapat membantu memetakan dan memvisualisasikan topik yang menjadi perbincangan dari masing-masing kubu pendukung capres & cawapres selama masa kampanye Pemilu 2019.

## 1.2 Rumusan Masalah

Perumusan masalah yang diangkat pada tugas akhir ini adalah:

1. Bagaimana mengakuisisi data akun beserta tweet dari calon legislatif pada media sosial Twitter?
2. Bagaimana melakukan pengolahan pada data sehingga bisa dilakukan permodelan topik dari tweet yang sudah dikumpulkan dari akun calon legislatif?

3. Bagaimana cara memvisualisasikan topik yang sudah didapatkan pada sebuah aplikasi?

### **1.3 Batasan Permasalahan**

Batasan permasalahan dalam tugas akhir ini adalah:

1. Pada setiap kubu capres & cawapres, dipilih masing-masing 15 akun caleg pada setiap partai yang mengikuti Pemilu 2019.
2. Data yang akan diambil pada penelitian ini hanya pada *tweet* dan *retweet* yang dibagikan oleh caleg tersebut, tanpa melibatkan balasan (*reply*) dari akun lain.
3. Data berupa *tweet* yang akan diolah diambil mulai dari tahun 2018, atau pada saat deklarasi capres & cawapres dilakukan.
4. Pembuatan model yang dilakukan menggunakan Author-Topic Model dilakukan sebatas untuk mengetahui topik yang dibahas pada oleh caleg pada setiap kubu pendukung capres & cawapres.

### **1.4 Tujuan**

Tujuan dari penelitian ini adalah memodelkan topik yang dibahas oleh caleg masing-masing kubu pendukung capres & cawapres, yang kemudian divisualisasikan dalam sebuah aplikasi, yang berguna untuk membantu mengetahui topik apakah yang menjadi fokus atau paling sering dibahas oleh para caleg dari masing-masing kubu capres & cawapres. Dengan adanya aplikasi ini, diharapkan dapat membantu masyarakat memantau topik kampanye apa saja yang disuarakan oleh para caleg di Twitter.

### **1.5 Manfaat**

Manfaat yang diperoleh dengan dilakukan penelitian ini adalah:

1. Membantu masyarakat dalam mengetahui topik kampanye yang dibagikan oleh para caleg.



2. Membantu melakukan analisis secara umum terkait *tweet* yang dibagikan oleh caleg selama Pemilu 2019.
3. Menyediakan data yang dapat digunakan sebagai acuan untuk menganalisis perkembangan kampanye politik yang dilakukan selama Pemilu 2019.

### **1.6 Relevansi**

Tugas akhir yang diusulkan berkaitan dengan mata kuliah Analisis dan Desain Perangkat Lunak, Sistem Cerdas, Konstruksi & Pengujian Perangkat Lunak, Pemrograman Berbasis Web, Desain Basis Data, dan Pengolahan Bahasa Alami.

*Halaman sengaja dikosongkan.*

## BAB II TINJAUAN PUSTAKA

Bab tinjauan pustaka terdiri dari landasan-landasan yang akan digunakan dalam penelitian tugas akhir ini, mencakup penelitian-penelitian sebelumnya, kajian pustaka, dan metode yang digunakan selama pengerjaan.

### 2.1 Penelitian Terkait

Berikut adalah penelitian yang terkait dengan tugas akhir yang diusulkan

**Tabel 2.1 Literatur 1**

<b>Judul</b>	Rancang Bangun Aplikasi Teenstagram untuk Mengelompokkan Topik Caption Akun Instagram dengan Permodelan Author Topic Models (Studi Kasus: Siswa SMA di Surabaya)[8]
<b>Penulis</b>	Faiz Nur Fitrah Insani
<b>Ringkasan</b>	Penelitian ini berfokus untuk merancang sebuah aplikasi berupa website yang bertujuan untuk melakukan visualisasi yang dapat mengetahui topik apa yang dibahas pada akun media sosial Instagram siswa SMA yang ada di Surabaya. Dengan mengumpulkan setiap akun siswa dan dikelompokkan sesuai dengan SMA masing-masing, kemudian dilakukan pengambilan <i>caption</i> atau uraian singkat pada setiap foto yang diunggah di Instagram. Setiap <i>caption</i> tersebut selanjutnya akan dilakukan <i>data pre-processing</i> , dan selanjutnya dilakukan pemrosesan data menggunakan Author-Topic Modeling untuk mengetahui topik apa saja yang paling banyak dibahas oleh siswa SMA.
<b>Keterkaitan penelitian</b>	Penelitian sebelumnya menganalisis <i>caption</i> unggahan foto Instagram siswa SMA di Surabaya, sementara tugas akhir yang diusulkan akan melakukan analisis pada setiap <i>tweet</i> dari akun

	Twitter calon legislatif pada setiap partai yang menjadi pendukung dari para capres & cawapres 2019.
--	--

Tabel 2.2 Literatur 2

<b>Judul</b>	Mining Latent Research Groups within Institutions Using an Author-Topic Model[9]
<b>Penulis</b>	Takeshi Shirai, Yoichi Tomiura, Shosaku Tanaka, dan Ryutaro Ono
<b>Ringkasan</b>	Penelitian ini bertujuan untuk membuat sebuah sistem pendukung pada sebuah administrator penelitian yang mampu mengidentifikasi <i>latent research group</i> pada sebuah institusi. Yang dimaksud <i>latent research group</i> adalah sepasang peneliti yang berasal dari bidang yang berbeda, dan memiliki keterkaitan pada setiap elemen penelitian dari peneliti. Pada penelitian ini digunakan Author-Topic Model untuk mengetahui kemungkinan distribusi dari setiap topik pada setiap <i>author</i> dan juga kemungkinan distribusi kejadian dari setiap kata pada setiap topik untuk setiap topik untuk menganalisis studi dari peneliti.
<b>Keterkaitan penelitian</b>	Penelitian tersebut melakukan pengelompokkan peneliti berdasarkan data studi yang dilakukan dan juga keterkaitan penelitiannya, sementara tugas akhir yang diusulkan akan mengelompokkan topik berdasarkan data berupa tweet dan membandingkan antara kedua kubu capres & cawapres 2019.

Tabel 2.3 Literatur 3

<b>Judul</b>	Author Topic Model-Based Collaborative Filtering for Personalized POI Recommendations[10]
<b>Penulis</b>	Shuhui Jiang, Xueming Qian, Jialie Shen, Yun Fu, and Tao Mei

<b>Ringkasan</b>	Penelitian ini bertujuan untuk memberikan personalisasi rekomendasi tujuan wisata dengan menggunakan metode author topic model-based collaborative filtering (ATCF). Preferensi topik dari user didapatkan dari deksripsi tekstual yang ada pada setiap foto yang sudah diunggah di situs Flickr melalui ATM, yang kemudian dapat memunculkan topik mengenai tujuan wisata dan preferensi dari user secara bersamaan. Selanjutnya Point of Interest atau POI dapat diketahui dengan melakukan penggolongan dari pengguna yang memiliki kemiripan pada preferensi tujuan wisata daripada menggunakan geo-tag GPS.
<b>Keterkaitan penelitian</b>	Penelitian tersebut bertujuan untuk memberikan rekomendasi tujuan wisata yang lebih baik, sementara tugas akhir yang diusulkan lebih mengarah kepada melakukan visualisasi topik dari setiap tweet yang dibagikan oleh setiap caleg sebagai pendukung dari masing-masing capres & cawapres pemilu 2019.

## 2.2 Dasar Teori

### 2.2.1 Twitter

Twitter adalah layanan jejaring sosial online yang memungkinkan pengguna untuk mengirim dan membaca pesan pendek yang disebut "tweets". Secara teknis, data Twitter dapat dibagi menjadi dua bagian, yaitu, data terstruktur seperti "id" dan "lokasi", dan data tidak terstruktur yang meliputi pesan teks, tautan pendek, dan sebagainya. Selain itu, tweeter atau pengguna Twitter memiliki beberapa perilaku sosial, seperti *follow* (memperhatikan orang lain), *mention* (“@” orang lain) dan *retweet* (meneruskan tweet orang lain)[11].

### 2.2.2 Pemilihan Umum 2019

Menurut Undang-Undang no. 7 tahun 2017 tentang Pemilihan Umum, Pemilihan Umum yang selanjutnya disebut Pemilu adalah sarana kedaulatan rakyat untuk memilih anggota Dewan Perwakilan Rakyat, anggota Dewan Perwakilan Daerah,

Presiden dan Wakil Presiden, dan untuk memilih anggota Dewan Perwakilan Rakyat Daerah, yang dilaksanakan secara langsung, umum, bebas, rahasia, jujur, dan adil dalam Negara Kesatuan Republik Indonesia berdasarkan Pancasila dan Undang-Undang Dasar Negara Republik Indonesia Tahun 1945 [2]. Setiap penyelenggaraan pemilu akan diramaikan dengan hadirnya partai-partai politik yang bertarung mendapatkan suara rakyat. Partai politik (*parties*), yaitu kumpulan individu dengan pemikiran yang sama, yang memiliki pemahaman organisasi dan struktur ideologi yang telah disepakati, untuk mencapai tujuan yang sama. Bagi partai politik, salah satu hal yang menjadi fokus utama agar dapat mencapai tujuan tersebut tergantung pada kemampuan mereka untuk berkomunikasi dengan mereka yang akan memilih dan melegitimasi mereka (*citizens*), dan tentunya partai politik memerlukan media dalam berkomunikasi, seperti melalui media cetak, maupun melalui cara *online* dengan menggunakan media sosial seperti Facebook atau Twitter[12].

Pada penyelenggaraan Pemilu 2019, total terdapat 16 partai politik nasional, dan 4 partai politik lokal di Aceh yang menjadi peserta Pemilu 2019. Jumlah ini bertambah jika dibandingkan dengan Pemilu Legislatif 2014 yang diikuti sebanyak 12 partai politik nasional dan 3 partai politik lokal[13]. Partai politik yang berpartisipasi dalam Pemilu 2019 antara lain:

1. Partai Kebangkitan Bangsa (PKB)
2. Partai Gerakan Indonesia Raya (Gerindra)
3. Partai Demokrasi Indonesia Perjuangan (PDIP)
4. Partai Golongan Karya (Golkar)
5. Partai Nasdem
6. Partai Gerakan Perubahan Indonesia (Garuda)
7. Partai Berkarya
8. Partai Keadilan Sejahtera (PKS)
9. Partai Persatuan Indonesia (Perindo)
10. Partai Persatuan Pembangunan (PPP)
11. Partai Solidaritas Indonesia (PSI)
12. Partai Amanat Nasional (PAN)
13. Partai Hati Nurani Rakyat (Hanura)

14. Partai Demokrat
15. Partai Bulan Bintang (PBB)
16. Partai Keadilan dan Persatuan Indonesia (PKPI)

Dari partai-partai tersebut, kemudian dapat dikelompokkan berdasarkan kubu pendukung calon presiden dan calon wakil presiden dalam tabel berikut[14]:

**Tabel 2.4 Pembagian Partai berdasarkan Kubu Pendukung [14]**

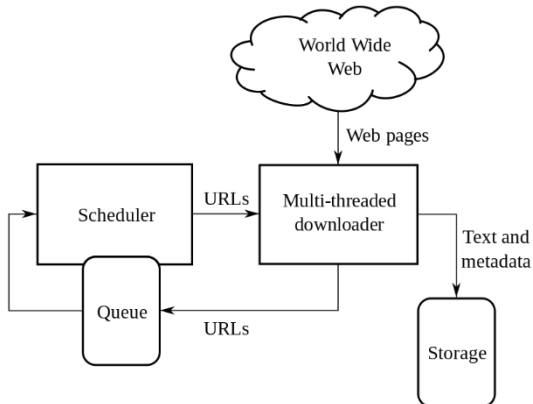
<b>Kubu Jokowi – Ma’ruf</b>	<b>Kubu Prabowo - Sandi</b>
PDIP, PPP, PKB, Golkar, Nasdem, PSI, Perindo, PKPI, dan Hanura	Gerindra, PKS, PAN, Demokrat, Partai Berkarya

Partai-partai yang sudah terdaftar dengan kurang lebih 500 calon legislatif DPR RI pada setiap partai akan berebut suara rakyat pada 80 dapil[15]. Salah satu cara untuk mendapatkan dukungan adalah dengan menggunakan media sosial, salah satunya Twitter. Anggota partai politik atau calon legislatif dapat memberikan opini yang kemudian menarik perhatian orang untuk mendukung calon legislatif yang juga berpengaruh pada perolehan suara dari partai tersebut. Suara dari partai politik tersebut selanjutnya dapat berpengaruh pada perolehan suara yang akan diperoleh oleh calon presiden & calon wakil presiden dalam suatu pemilihan umum.

Bagi para calon legislatif, tentunya penting untuk tetap menyampaikan gagasan yang sesuai dengan tujuan dari partai yang mengusungnya. Pada penelitian ini, salah satu yang menjadi fokus utama adalah untuk mengetahui apakah antara satu partai politik memiliki kesamaan pendapat atau topik bahasan pada setiap *tweet* yang dibagikan melalui akun Twitter masing-masing calon legislatif atau pendukung salah satu kubu calon presiden dan calon wakil presiden, karena dengan memiliki kesamaan gagasan atau topik yang dibahas dapat memperkuat tujuan utama yang ingin dicapai oleh suatu koalisi dari para pendukung calon presiden & calon wakil presiden.

### 2.2.3 Social Media Crawler

Media sosial adalah sebuah teknologi media komunikasi interaktif yang memungkinkan penggunanya untuk bisa membagikan informasi, ide, minat, ataupun bentuk ekspresi melalui suatu komunitas virtual dan jaringan[16]. Konten dari media sosial dapat dikategorikan berdasarkan beberapa tingkatan, yaitu informasi, hiburan, *money-making*, dan untuk interaksi sosial[17]. Sedangkan yang dimaksud *web crawler* atau yang bisa disebut *robots*, *spiders*, *worms*, *walkers*, dan *wanderers* digunakan untuk melakukan penelusuran secara otomatis untuk memberikan indeks pada website[18].



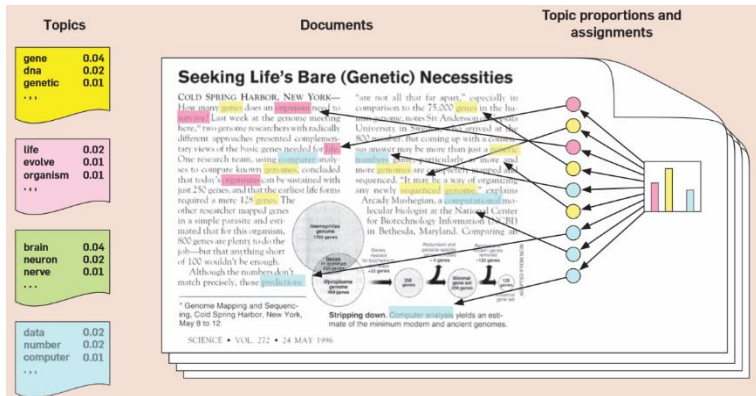
**Gambar 2.1** Arsitektur Web Crawler

Sementara yang dimaksud dengan *social media crawler* adalah menggunakan *web crawler* pada media sosial, khususnya Twitter. Data yang akan digunakan adalah *tweet* dari berbagai akun Twitter yang termasuk salah satu contoh media sosial dengan banyak pengguna, yang kemudian dikumpulkan dengan menggunakan bantuan dari *web crawler* untuk kemudian dilakukan pemrosesan data lebih lanjut.



## 2.2.4 Topic Model

Topic model adalah algoritma yang digunakan untuk menemukan tema utama pada sebuah dokumen tidak terstruktur yang berukuran besar, dan juga dapat mengatur koleksi sesuai dengan tema yang ditemukan. Selain itu, topic model dapat disesuaikan dengan berbagai jenis data, dan dapat digunakan untuk menemukan pola dalam data genetik, gambar, dan jejaring sosial[19].



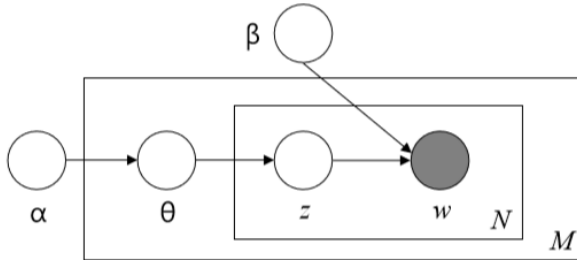
Gambar 2.2 Konsep Permodelan Topik[19]

Gambar 2.2 merupakan contoh penerapan dari topic model. Di sebelah kiri, asumsikan bahwa beberapa topik yang tersebar pada beberapa kata, sudah terdapat di seluruh koleksi. Setiap dokumen diasumsikan akan dibuat dengan cara berikut. Pertama, pilih sebuah distribusi dari topik (ditunjukkan dengan histogram pada sebelah kanan), lalu untuk setiap kata, tetapkan sebuah topik (ditunjukkan dengan lingkaran berwarna), dan pilih kata dari topik yang berhubungan[19].

## 2.2.5 Latent Dirichlet Allocation

Latent Dirichlet Allocation (LDA) adalah sebuah model generatif yang membahas kemungkinan pada sebuah korpus. LDA adalah sebuah model Bayesian dengan memiliki 3

tingkatan hirarki, yang pada setiap item pada suatu kumpulan dimodelkan sebagai suatu campuran terbatas atas serangkaian topik yang mendasarinya[20].

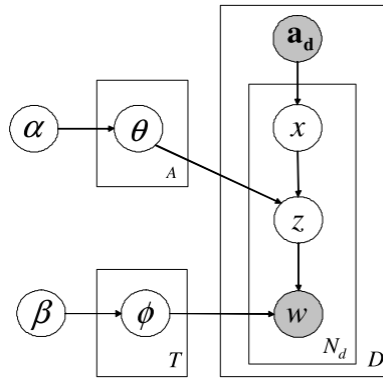


**Gambar 2.3 Konsep Latent Dirichlet Allocation[20]**

Gambar 2.3 menunjukkan model LDA, dimana pada gambar tersebut menunjukkan tiga tingkatan pada representasi LDA. Parameter  $\alpha$  dan  $\beta$  adalah parameter pada tingkatan corpus, diasumsikan untuk diuji sekali pada proses pembuatan sebuah korpus. Variabel  $\theta$  adalah variabel pada tingkatan dokumen, diuji sekali pada setiap dokumen. Variable  $z$  dan  $w$  adalah pada tingkatan kata atau *word* yang diujikan sekali pada setiap kata di setiap dokumen[20].

### 2.2.6 Author-Topic Model

Author-Topic Model adalah salah satu pengembangan dari Latent Latent Dirichlet Allocation (LDA) untuk memuat informasi mengenai penulis atau *author*. Setiap penulis diasosiasikan dengan sebuah distribusi multinomial pada beberapa topik, dan setiap topik berasosiasi dengan sebuah distribusi multinomial pada beberapa kata. Sebuah dokumen dengan beberapa penulis dimodelkan sebagai sebuah distribusi pada beberapa topik yang merupakan campuran dari distribusi yang dikaitkan dengan penulis[21].



**Gambar 2.4 Konsep Author Topic Model[21]**

Gambar 2.4 menunjukkan konsep dari Author-Topic Model. Seperti yang ada pada Author Model (LDA),  $x$  menandakan *author* yang berkaitan dengan suatu kata yang diberikan, dipilih dari  $a_d$ , memutuskan untuk menulis dokumen  $d$ . Untuk setiap kata yang berasosiasi dengan distribusi pada beberapa topik,  $\theta$ , dipilih dari Dirichlet asimetris ( $\alpha$ ) sebelumnya. Bobot campuran sesuai dengan penulis yang dipilih digunakan untuk memilih topik  $z$ , dan kata dihasilkan sesuai dengan distribusi  $\phi$  sesuai dengan topik itu, diambil dari Dirichlet ( $\beta$ ) simetris sebelumnya. Dengan mengestimasi parameter  $\phi$  dan  $\theta$ , didapatkan informasi mengenai topik mana dan author mana yang ditulis, yang juga ditunjukkan dengan konten pada setiap dokumen dalam hal topik-topik tersebut.

TOPIC 19		TOPIC 24		TOPIC 29		TOPIC 87	
WORD	PROB.	WORD	PROB.	WORD	PROB.	WORD	PROB.
LIKELIHOOD	0.0539	RECOGNITION	0.0400	REINFORCEMENT	0.0411	KERNEL	0.0683
MIXTURE	0.0509	CHARACTER	0.0336	POLICY	0.0371	SUPPORT	0.0377
EM	0.0470	CHARACTERS	0.0250	ACTION	0.0332	VECTOR	0.0257
DENSITY	0.0398	TANGENT	0.0241	OPTIMAL	0.0208	KERNELS	0.0217
GAUSSIAN	0.0349	HANDWRITTEN	0.0169	ACTIONS	0.0208	SET	0.0205
ESTIMATION	0.0314	DIGITS	0.0159	FUNCTION	0.0178	SYM	0.0204
LOG	0.0263	IMAGE	0.0157	REWARD	0.0165	SPACE	0.0188
MAXIMUM	0.0254	DISTANCE	0.0153	SUTTON	0.0164	MACHINES	0.0168
PARAMETERS	0.0209	DIGIT	0.0149	AGENT	0.0136	REGRESSION	0.0155
ESTIMATE	0.0204	HAND	0.0126	DECISION	0.0118	MARGIN	0.0151
AUTHOR	PROB.	AUTHOR	PROB.	AUTHOR	PROB.	AUTHOR	PROB.
Tresp_V	0.0333	Simard_P	0.0694	Singh_S	0.1412	Smola_A	0.1033
Singer_Y	0.0281	Martin_G	0.0394	Barlo_A	0.0471	Scholkopf_B	0.0730
Jebara_T	0.0207	LeCun_Y	0.0359	Sutton_R	0.0430	Burges_C	0.0489
Ghahramani_Z	0.0196	Denker_J	0.0278	Dayan_P	0.0324	Vapnik_V	0.0431
Ueda_N	0.0170	Henderson_D	0.0256	Parr_R	0.0314	Chapelle_O	0.0210
Jordan_M	0.0150	Revow_M	0.0229	Dietterich_T	0.0231	Cristianini_N	0.0185
Roweis_S	0.0123	Platt_J	0.0226	Tsitsiklis_J	0.0194	Ratsch_G	0.0172
Schuster_M	0.0104	Keeler_J	0.0192	Randlov_J	0.0167	Laskov_P	0.0169
Xu_L	0.0098	Rashid_M	0.0182	Bradtko_S	0.0161	Tipping_M	0.0153
Saul_L	0.0094	Sackinger_E	0.0132	Schwartz_A	0.0142	Sollich_P	0.0141

**Gambar 2.5** Contoh Topik dari Model Author-Topic dengan dataset Neural Information Processing Systems (NIPS) Conference berdasarkan topik[21]

PROB. TOPIC		AUTHOR = Jordan_M	
		WORDS	
.1389	37	MIXTURE, EM, LIKELIHOOD, EXPERTS, MIXTURES, EXPERT, GATING, PARAMETERS, LOG, JORDAN	
.1221	60	BELIEF, FIELD, STATE, APPROXIMATION, MODELS, VARIABLES, FACTOR, JORDAN, NETWORKS, PARAMETERS	
.0598	52	ALGORITHM, ALGORITHMS, PROBLEM, STEP, PROBLEMS, LINEAR, UPDATE, FIND, LINE, ITERATIONS	
.0449	77	MOTOR, TRAJECTORY, ARM, INVERSE, HAND, CONTROL, MOVEMENT, JOINT, DYNAMICS, FORWARD	

PROB. TOPIC		AUTHOR = Koch_C	
		WORDS	
.2518	4	LIGHT, RESPONSE, INTENSITY, RETINA, OPTICAL, KOCH, BACKGROUND, CONTRAST, CENTER, FEEDBACK	
.0992	45	VISUAL, STIMULUS, CORTEX, SPATIAL, ORIENTATION, RESPONSE, CORTICAL, RECEPTIVE, TUNING, STIMULI	
.0882	84	SPIKE, FIRING, SYNAPTIC, SYNAPSES, MEMBRANE, POTENTIAL, CURRENT, SPIKES, RATE, SYNAPSE	
.0504	64	CIRCUIT, CURRENT, VOLTAGE, ANALOG, CHIP, VLSI, CIRCUITS, SILICON, PULSE, MEAD	

PROB. TOPIC		AUTHOR = LeCun_Y	
		WORDS	
.2298	13	RECOGNITION, CHARACTER, TANGENT, CHARACTERS, DISTANCE, HANDWRITTEN, DIGITS, SEGMENTATION, DIGIT, IMAGE	
.0930	53	GRADIENT, FUNCTION, DESCENT, ERROR, VECTOR, DERIVATIVE, DERIVATIVES, OPTIMIZATION, PARAMETERS, LOCAL	
.0930	69	LAYER, WEIGHTS, PROPAGATION, BACK, OUTPUT, LAYERS, INPUT, NUMBER, WEIGHT, FORWARD	
.0762	36	INPUT, OUTPUT, INPUTS, OUTPUTS, VALUES, ARCHITECTURE, SUM, ADAPTIVE, PREVIOUS, PROCESSING	

**Gambar 2.6** Contoh Topik dari Model Author-Topic dengan dataset Neural Information Processing Systems (NIPS) Conference berdasarkan author[21]

Contoh dari penerapan ATM adalah dengan menggunakan data set dari konferensi NIPS. Pada Gambar 2.5, menunjukkan 8 dari 100 topik yang dipelajari oleh model dari korpus NIPS[21]. Topik-topik tersebut didapatkan dengan menghitung kemungkinan dari setiap kata dan *author* yang terbanyak.

## 2.2.7 Perplexity

Perplexity adalah sebuah standar pengukuran untuk memperkirakan performa dari probabilistik sebuah model.

$$\text{perplexity}(\mathbf{w}_d | \mathbf{a}_d) = \exp \left[ -\frac{\ln p(\mathbf{w}_d | \mathbf{a}_d)}{N_d} \right]$$

Semakin rendah angka *perplexity* yang ada pada suatu dokumen, maka secara general semakin baik juga performanya[21]. Perplexity digunakan untuk melakukan validasi pada model yang sudah dibuat, sehingga pada hasil akhirnya didapatkan model yang terbaik.

## 2.2.8 Tableau

atm_hasitopik	atm_prob_topik	atm_prob_kata_topik	atm_prob_topik_id_author	atm_prob_topik_prob_author	atm_prob_topik_kategori	atm_prob_topik_partai	atm_prob_topik_id_kubu	atm_prob_topik_nama_kubu	atm_prob_topik_avg
0	0.00	perdaye	102	0.99	Yes	Hukum	PDIP	01	Jakarta-Ma'nu'
0	0.00	perdaye	102	0.99	Yes	Hukum	PDIP	01	Jakarta-Ma'nu'
0	0.00	perdaye	102	0.99	Yes	Hukum	PDIP	01	Jakarta-Ma'nu'
0	0.00	perdaye	102	0.99	Yes	Hukum	PDIP	01	Jakarta-Ma'nu'
0	0.00	perdaye	102	0.99	Yes	Hukum	PDIP	01	Jakarta-Ma'nu'
0	0.00	perdaye	102	0.99	Yes	Hukum	PDIP	01	Jakarta-Ma'nu'
0	0.00	perdaye	102	0.99	Yes	Hukum	PDIP	01	Jakarta-Ma'nu'
0	0.00	perdaye	102	0.99	Yes	Hukum	PDIP	01	Jakarta-Ma'nu'

**Gambar 2.7 Contoh Tampilan Tableau Desktop**

Tableau sebuah perangkat lunak yang dapat membantu proses visualisasi data yang banyak digunakan pada industri kecerdasan bisnis. Tableau membantu mengubah data yang masih mentah menjadi bentuk yang lebih mudah dipahami dan divisualisasikan. Beberapa fitur dari Tableau yang berguna antara lain dalam hal pencampuran data, analisis secara *real time*, dan dalam hal kolaborasi antar data. Selain itu untuk mengoperasikan Tableau tidak memerlukan keahlian teknis dalam hal pemrograman untuk mengoperasikan.

Beberapa produk Tableau yang digunakan pada penelitian ini adalah Tableau Desktop, yaitu perangkat lunak Tableau yang dipasang pada komputer pribadi dan digunakan untuk mengolah data secara personal[22]. Untuk mengakses Tableau Desktop, perlu didapatkan lisensi untuk bisa menggunakan produk secara keseluruhan. Lisensi ini bisa didapatkan dengan mendaftarkan diri pada situs Tableau sebagai *student* dan memasukkan data terkait dengan status sebagai mahasiswa.

Setelah dilakukan pengolahan data pada Tableau Desktop, untuk memudahkan dalam mengakses secara luas *workbook* yang sudah dibuat, *workbook* tersebut dapat disimpan pada akun Tableau Public yang sudah dibuat sebelumnya. Dengan bantuan Tableau Public, setiap *workbook* yang sudah dibuat secara *local* dapat diakses secara *public* dan dapat dibagikan secara *online*. Contoh tampilan Tableau Public dapat dilihat pada Gambar 2.8.



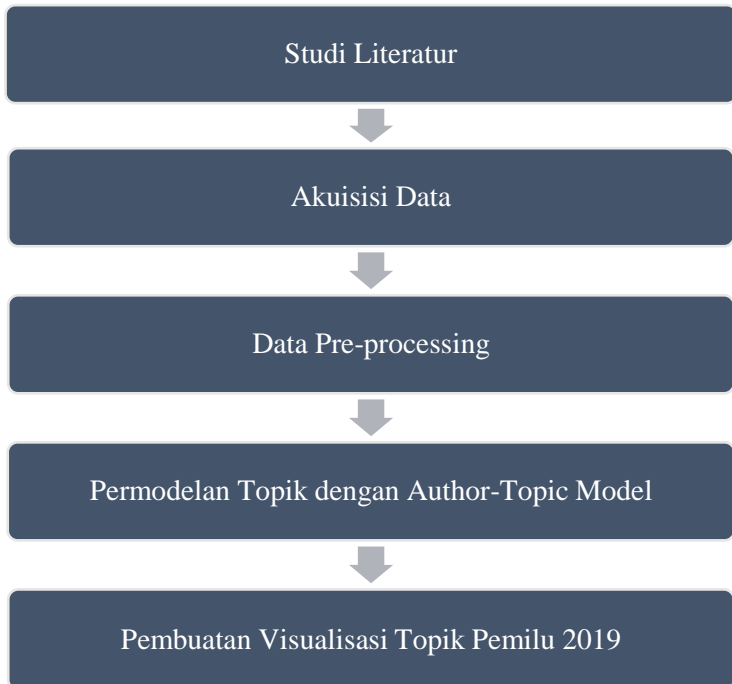
**Gambar 2.8 Contoh Tampilan Tableau Public**

## **BAB III METODOLOGI**

Pada bab ini diuraikan beberapa tahapan metodologi yang perlu dilakukan dalam tugas akhir ini.

### **3.1 Tahapan Pelaksanaan Tugas Akhir**

Pada sub bab ini akan menjelaskan mengenai metodologi dalam pengerjaan tugas akhir. Metodologi dapat dilihat pada Gambar 3.1:



**Gambar 3.1 Metodologi Penelitian**

### 3.1.1 Studi Literatur

Pada proses studi literatur dilakukan pencarian terhadap materi-materi yang terkait dengan penelitian tugas akhir. Materi-materi yang dicari meliputi:

1. Konsep Social Media Analysis
2. Konsep pengambilan data dari media sosial
3. Konsep Author-Topic Model
4. Konsep pembuatan visualisasi topik Pemilu 2019

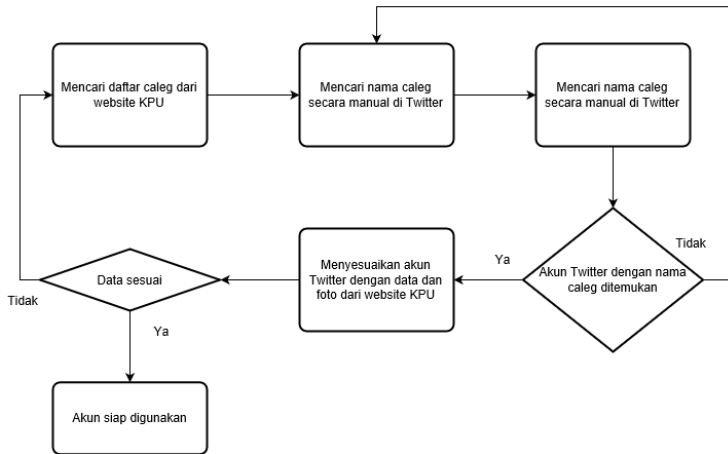
Luaran yang dihasilkan dari proses studi literatur ini adalah mengetahui adanya *knowledge gap* dan pemahaman yang dibutuhkan berdasarkan literatur yang sudah dikumpulkan.

### 3.1.2 Akuisisi Data

Proses akuisisi untuk mendapatkan akun Twitter dari 2 kubu calon presiden dilakukan dengan beberapa tahapan, yaitu:

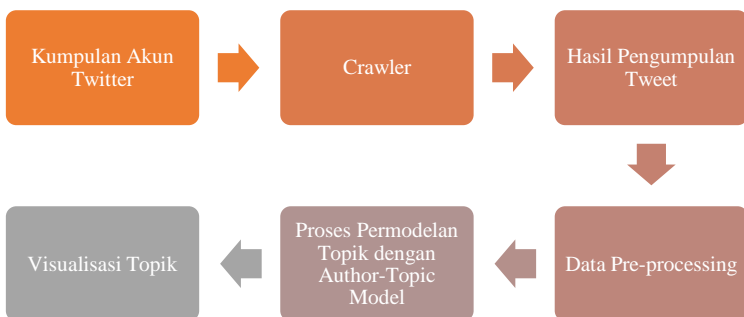
- Mencari akun Twitter dari para calon anggota legislatif ataupun pendukung tiap kubu calon presiden. Proses pencarian akan dilakukan secara manual dengan menggunakan data daftar calon legislatif pada website Komisi Pemilihan Umum.
- Melakukan validasi pada setiap akun Twitter calon anggota legislatif ataupun pendukung tiap kubu calon presiden, hal ini dilakukan untuk mengetahui apakah akun tersebut terverifikasi dan masih aktif digunakan. Selain itu validasi juga dilakukan apakah akun tersebut merupakan akun dari caleg yang dimaksud dan sesuai dengan data dari KPU atau akun orang lain. Proses validasi ditunjukkan pada Gambar 3.2:





**Gambar 3.2 Proses Validasi Akun Twitter Caleg**

- Mengumpulkan setiap *tweet* yang dibagikan oleh masing-masing akun calon anggota legislatif tersebut. Proses pengumpulan *tweet* dilakukan dengan melakukan crawling pada setiap akun yang sudah didapatkan dan kemudian menyimpan kumpulan *tweet* tersebut. Secara umum, proses pengumpulan data dilakukan seperti pada gambar berikut:



**Gambar 3.3 Proses Pengumpulan data status Twitter**

### 3.1.3 Data Pre-processing

Proses ini dilakukan untuk mengolah data berupa *tweet* yang sudah dikumpulkan sebelumnya. Beberapa tahapan yang dilakukan dalam data pre-processing antara lain:

#### 3.1.1.1 Case Folding

*Case folding* adalah tahapan untuk menyeragamkan bentuk teks menjadi *lowercase* atau *uppercase*, sehingga teks yang akan diolah memiliki bentuk yang sama. Contoh dari *case folding* menjadi *lowercase* adalah seperti berikut:

**Tabel 3.1 Contoh Case Folding**

Sebelum <i>Case Folding</i>	Setelah <i>Case Folding</i>
Selamat pagi. Mari menyambut awal pekan ini dengan optimisme dan kesadaran bahwa bangsa ini, Indonesia ini, adalah negara besar. Indonesia juga dianugerahi keragaman dan kemajemukan suku bangsa, agama, adat, tradisi, hingga bahasa daerah. Semua perbedaan itu sudah menjadi hukum Allah.	selamat pagi. mari menyambut awal pekan ini dengan optimisme dan kesadaran bahwa bangsa ini, indonesia ini, adalah negara besar. indonesia juga dianugerahi keragaman dan kemajemukan suku bangsa, agama, adat, tradisi, hingga bahasa daerah. semua perbedaan itu sudah menjadi hukum allah.

### 3.1.3.2 Stemming

Stemming adalah tahapan untuk mengganti bentuk dari kata menjadi kata dasar, dengan menghilangkan imbuhan dari setiap kata tersebut. Jenis imbuhan yang dihilangkan antara lain awalan, sisipan, akhiran, dan kombinasi awalan dan akhiran. Pengubahan kata menjadi kata dasar ini dilakukan untuk menyamakan arti pada tiap kata yang bisa bermakna berbeda karena adanya imbuhan pada kata tersebut. Stemming dilakukan dengan menggunakan *library* Sastrawi, yaitu *library stemmer*.

**Tabel 3.2 Contoh Stemming**

Sebelum <i>Stemming</i>		Setelah <i>Stemming</i>	
'selamat',	'menyambut',	'selamat',	'sambut', 'pekan',
'pekan',	'optimisme',	'optimisme',	'sadar',
'kesadaran',	'bangsa',	'bangsa',	'indonesia',
'indonesia',	'negara',	'negara',	'indonesia',
'indonesia',	'dianugerahi',	'anugerah',	'ragam',
'keragaman',	'kemajemukan',	'majemuk',	'suku', 'bangsa',
'suku',	'bangsa', 'agama',	'agama',	'adat', 'tradisi',
'adat',	'tradisi', 'bahasa',	'bahasa',	'daerah', 'beda',
'daerah',	'perbedaan',	'hukum',	'allah'
'hukum',	'allah'		

### 3.1.3.3 Tokenization

Tokenization adalah tahapan untuk memisahkan setiap kalimat yang sudah didapatkan dalam bentuk setiap kata yang terpisah. Proses *tokenization* digunakan untuk mempermudah proses *stopwords removal* dan *stemming*. Contoh proses *tokenization* adalah sebagai berikut:

**Tabel 3.3 Contoh Tokenization**

Sebelum <i>Tokenization</i>	Setelah <i>Tokenization</i>
selamat pagi. mari menyambut awal pekan ini dengan optimisme dan kesadaran bahwa bangsa ini, indonesia ini, adalah negara besar. indonesia juga dianugerahi keragaman dan kemajemukan suku bangsa, agama, adat, tradisi, hingga bahasa daerah. semua perbedaan itu sudah menjadi hukum allah.	'selamat', 'pagi', 'mari', 'menyambut', 'awal', 'pekan', 'ini', 'dengan', 'optimisme', 'dan', 'kesadaran', 'bahwa', 'bangsa', 'ini', 'indonesia', 'ini', 'adalah', 'negara', 'besar', 'indonesia', 'juga', 'dianugerahi', 'keragaman', 'dan', 'kemajemukan', 'suku', 'bangsa', 'agama', 'adat', 'tradisi', 'hingga', 'bahasa', 'daerah', 'semua', 'perbedaan', 'itu', 'sudah', 'menjadi', 'hukum', 'allah'

### 3.1.3.4 Stopwords Removal

Stopwords removal adalah proses untuk menghapus kata yang sering digunakan atau muncul dalam jumlah yang banyak. Stopwords atau common words biasanya dianggap tidak memiliki makna, oleh karena itu diperlukan adanya penghapusan stopwords agar tidak mempengaruhi interpretasi dari data yang sudah dikumpulkan. Berikut adalah contoh penghapusan *stopwords*.

**Tabel 3.4 Contoh Stopwords Removal**

Sebelum <i>Stopwords Removal</i>	Setelah <i>Stopwords Removal</i>
'selamat', 'pagi', 'mari', 'menyambut', 'awal', 'pekan', 'ini', 'dengan', 'optimisme', 'dan',	'selamat', 'menyambut', 'pekan', 'optimisme', 'kesadaran', 'bangsa', 'indonesia', 'negara',

'kesadaran', 'bahwa', 'bangsa', 'ini', 'indonesia', 'ini', 'adalah', 'negara', 'besar', 'indonesia', 'juga', 'dianugerahi', 'keragaman', 'dan', 'kemajemukan', 'suku', 'bangsa', 'agama', 'adat', 'tradisi', 'hingga', 'bahasa', 'daerah', 'semua', 'perbedaan', 'itu', 'sudah', 'menjadi', 'hukum', 'allah'	'indonesia', 'dianugerahi', 'keragaman', 'kemajemukan', 'suku', 'bangsa', 'agama', 'adat', 'tradisi', 'bahasa', 'daerah', 'perbedaan', 'hukum', 'allah'
---	--

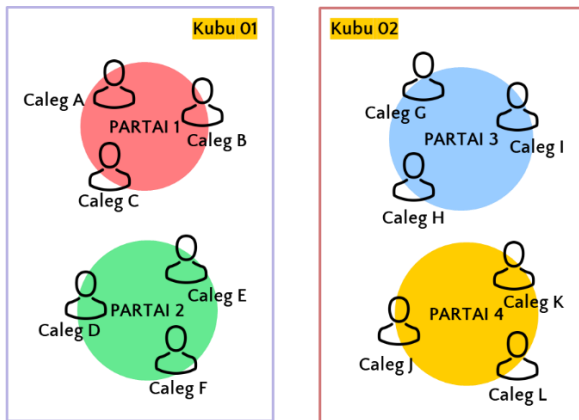
### 3.1.4 Permodelan Topik

Pembentukan model dimulai dengan memetakan setiap *tweet* secara acak pada beberapa topik dan *author* dari kumpulan *author* atau akun Twitter dan topik yang didapat. Pada mulanya, dari setiap akun Twitter masing-masing caleg akan dikumpulkan berdasarkan partai masing-masing, dan kemudian dikelompokkan menjadi 2 kubu pendukung calon presiden dan calon wakil presiden, yang kemudian dapat menunjukkan kecenderungan topik yang dibahas pada suatu partai tertentu, yang akan menunjukkan persebaran topik apa saja yang kemungkinan dibahas beserta kemungkinan siapakah *author* atau kubu yang membahas topik tersebut. Kemudian dilakukan iterasi sampai jumlah tertentu dengan tujuan untuk mendapatkan hasil terbaik. Setelah dilakukan iterasi, akan didapatkan nilai *perplexity* yang selanjutnya juga dapat digunakan untuk menentukan topik apa saja yang berhasil didapatkan dari pengolahan data tersebut.

### 3.1.5 Pembuatan Visualisasi Topik Pilpres 2019

Pada tahap ini dilakukan pembuatan aplikasi untuk memvisualisasikan hasil analisis permodelan topik yang sudah

dilakukan. Visualisasi akan dilakukan dengan membedakan pengelompokan topik berdasarkan partai pendukung dan juga berdasarkan kubu pendukung calon presiden & calon wakil presiden. Nantinya 1 partai akan dianggap sebagai 1 *author*, dan 1 kubu pendukung akan dianggap sebagai 1 *author*. Visualisasi akan dibagi berdasarkan beberapa kategori, yaitu topik yang dibicarakan dalam lingkup partai, dan dalam lingkup kubu pendukung.



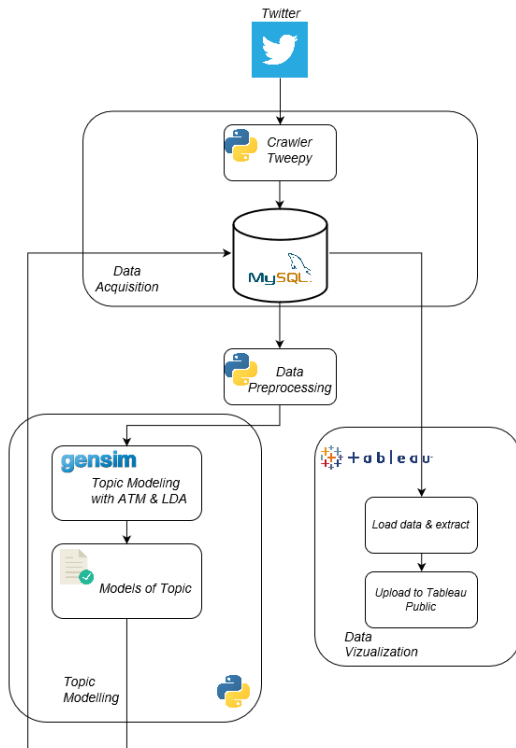
**Gambar 3.4 Pengelompokan Caleg untuk Permodelan Topik**

Setelah didapatkan model topik yang dibutuhkan, selanjutnya topik-topik tersebut akan divisualisasikan menggunakan Tableau. Untuk memudahkan pembuatan visualisasi, akan dibuat beberapa *workbook* sesuai dengan jenis data yang ingin divisualisasikan, baik itu data terkait analisis *tweet* secara umum, maupun analisis mengenai hasil permodelan topik yang sudah dilakukan. Setelah semua visualisasi sudah dibuat, selanjutnya akan diunggah ke akun Tableau Public yang sudah dibuat sebelumnya, agar hasil visualisasi dapat dibagikan dan dilihat secara *online*.

## BAB IV PERANCANGAN

Pada bab ini diuraikan perancangan terkait beberapa hal yang diperlukan dalam proses pembuatan visualisasi sesuai dengan alur yang sudah dijelaskan pada bab metodologi. Berikut adalah beberapa perancangan yang perlu dilakukan.

### 4.1 Arsitektur Sistem



**Gambar 4.1 Arsitektur Sistem**

Gambar 4.1 menunjukkan perancangan arsitektur sistem dilakukan untuk memberi gambaran mengenai sistem yang akan dibuat. Pada penelitian ini, akan dibuat sistem yang mampu

melakukan *crawling* data *tweet* dari akun caleg, kemudian melakukan pre-processing pada *tweet*, lalu *tweet* yang sudah dilakukan pre-processing diolah menggunakan library Gensim untuk melakukan permodelan topik. Hasil permodelan topik yang didapatkan, baik yang dibuat menggunakan metode ATM ataupun LDA tersebut kemudian ditampilkan menggunakan Tableau.

## 4.2 Pengambilan Data

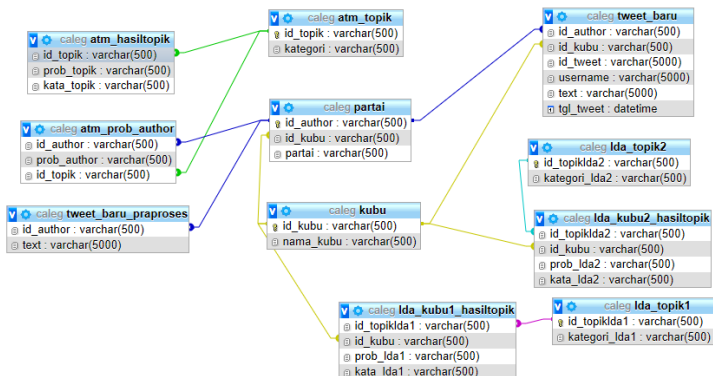
Dalam pembuatan visualisasi pengelompokkan topik status Twitter caleg, data yang diperlukan adalah kumpulan status Twitter caleg dari total 14 partai dari 2 kubu pendukung capres dan cawapres. Dari masing-masing partai, kemudian akan dipilih 20 akun Twitter caleg yang masih aktif dan sering memperbarui *tweetnya*. Kemudian akun-akun tersebut dikumpulkan menjadi satu kemudian diberi keterangan sesuai dengan partai yang mendukung caleg tersebut dan juga kubu capres & cawapres yang didukung, untuk mempermudah proses pengumpulan status Twitter dari masing-masing caleg.

Pengumpulan status Twitter dilakukan dengan menggunakan Tweepy, dengan ketentuan bahwa *tweet* yang diambil adalah *tweet* yang dibuat sejak 22 Mei 2018 sampai dengan dilakukannya pengumuman hasil Pemilu 2019 yaitu pada tanggal 22 Mei 2019. Tanggal ini dipilih karena untuk pengolahan dan pembentukan topik membutuhkan data yang cukup banyak agar hasilnya semakin baik dan representatif.

Selanjutnya setiap *tweet* yang berhasil dikumpulkan dimasukkan pada database “Caleg” yaitu pada tabel “*tweet\_baru*”. Beberapa informasi yang juga diambil selain *tweetnya* adalah *id\_tweet*, *username*, dan *timestamp* dari *tweet* tersebut. Dari hasil *crawling* tersebut, didapatkan total kurang lebih 132,838 *tweet* dari 14 partai pendukung dalam Pemilu 2019.



### 4.3 Desain Basis Data Penyimpanan Tweet dan Akun Twitter



Gambar 4.2 Desain Penyimpanan Basis Data

Pada penelitian ini, dibuat sebuah basis data “Caleg” untuk menyimpan semua data yang terkait. Dalam perancangan basis data terbagi menjadi beberapa tabel yang digunakan untuk menyimpan hasil crawling tweet dari caleg dan menyimpan hasil permodelan topik yang sudah dilakukan. Berikut adalah beberapa keterangan mengenai beberapa tabel yang ada pada basis data “Caleg”:

- Tabel “partai”, digunakan untuk menyimpan kode partai (id\_author) dan juga nama partai.
- Tabel “atm\_topik”, digunakan untuk menyimpan id\_topik dan juga kategori topik yang dihasilkan dalam permodelan menggunakan author-topic model.
- Tabel “tweet\_baru” digunakan untuk menyimpan hasil crawling tweet kedalam basis data.
- Tabel “tweet\_baru\_praproses” untuk menyimpan hasil tweet yang sebelumnya sudah dilakukan pre-processing.
- Tabel “atm\_hasitopik” berisi detail kata yang ada pada setiap topik beserta probabilitas dari setiap kata yang ada pada topik yang dihasilkan dari author-topic model.
- Tabel “atm\_prob\_author” berisi probabilitas author dalam setiap topik yang sudah dibuat sebelumnya.

- Tabel “atm\_topik” berisi id\_topik beserta masing-masing kategori topik yang didapatkan dari permodelan menggunakan ATM.
- Tabel “kubu” berisi id\_kubu dan nama\_kubu yang digunakan untuk permodelan topik dari masing-masing kubu
- Tabel “lda\_hasiltopik” berisi kata-kata yang menyusun suatu topik beserta probabilitasnya yang didapatkan dari permodelan menggunakan LDA dengan author masing-masing kubu.
- Tabel “lda\_topik” berisi id\_topiklda1 beserta kategori topik yang didapatkan setelah melakukan permodelan menggunakan LDA.

#### 4.4 Data Pre-Processing

Tahapan ini mencakup beberapa hal yang dilakukan untuk mempersiapkan data sebelum dilakukan pengolahan data. Beberapa tahapan yang dilakukan mulai dari case folding, stemming, stopword removal, dan tokenization. Penjelasan lebih detail ada pada Gambar 4.3:



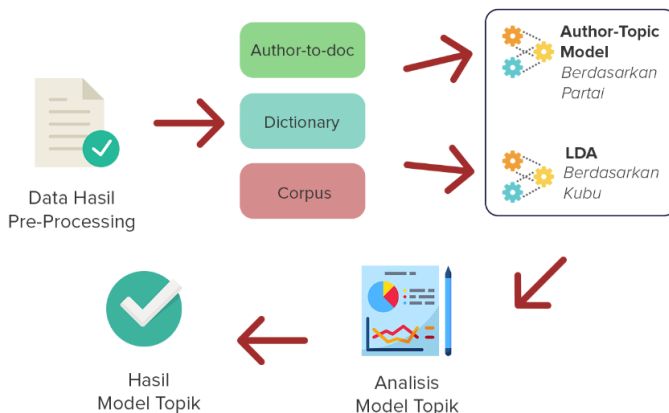
**Gambar 4.3 Pre-processing Tweet**

1. Cleansing adalah tahapan awal yang dilakukan untuk menghilangkan karakter-karakter tidak digunakan, seperti symbol, angka, emoji, dan menghapus beberapa baris yang kosong.
2. Case folding dilakukan untuk mengubah format huruf pada dokumen. Pada penelitian ini, semua huruf yang ada di dokumen diubah menjadi huruf kecil atau *lowercase*.
3. Stemming dilakukan untuk menghilangkan imbuhan pada kata, dan mengubah kata tersebut menjadi kata dasar.

4. Tokenization dilakukan untuk memisahkan *tweet* menjadi bagian per kata.
5. Stopwords removal dilakukan untuk menghapus kata-kata yang sering muncul atau sering digunakan tetapi tidak mempunyai nilai informasi. Pada kasus ini, dibuat daftar stopwords khusus yang banyak muncul selama masa Pemilu 2019.

## 4.5 Pemrosesan Data

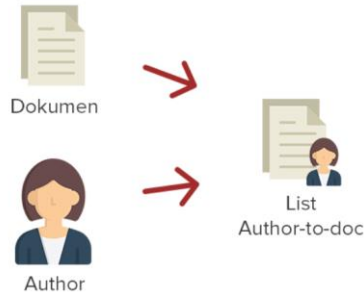
Pada tahapan pemrosesan data, data yang sebelumnya sudah dilakukan pre-processing. Gambar 4.4 menjelaskan bagaimana pemrosesan data dilakukan.



**Gambar 4.4** Gambaran Umum Pembuatan Model Topik

### 4.5.1 Pembentukan Author-to-doc

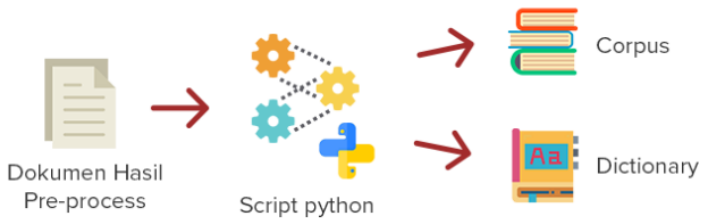
Proses pembentukan author-to-doc dilakukan dengan mendefinisikan terlebih dahulu author pada setiap dokumennya. Author-to-doc adalah daftar dokumen yang akan diolah dengan *author* atau penulis dari dokumen tersebut. Pada penelitian ini, dokumen yang dimaksud adalah *tweet* yang sudah dikumpulkan, dan *author*-nya adalah asal partai dari akun Twitter yang sudah dikumpulkan sebelumnya.



**Gambar 4.5 Proses Pembuatan Author-to-doc**

### 4.5.2 Pembentukan Dictionary dan Corpus

Pembentukan *dictionary* dan *corpus* dilakukan untuk melakukan *topic modeling*. Dictionary adalah format data yang berisi himpunan data unik yang memiliki indeks, yang dapat memudahkan saat menampilkan kata yang termasuk pada suatu model. Sedangkan *corpus* sendiri adalah format data yang memiliki bentuk dokumen term-matrix, yang selanjutnya akan digunakan untuk melakukan eksperimen saat pembentukan model. Gambar 4.6 menjelaskan proses pembentukan dictionary dan corpus.

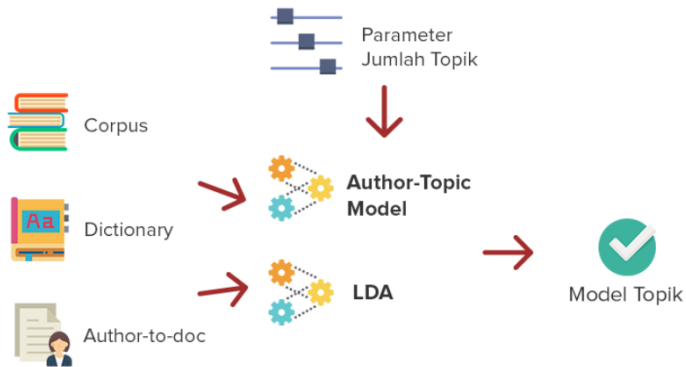


**Gambar 4.6 Proses Pembentukan Dictionary dan Corpus**

### 4.5.3 Pembentukan Topic Model

Proses pembentukan topic model dilakukan dengan menggunakan Author-Topic Model dan Latent Dirichlet Allocation. Penggunaan 2 metode ini dikarenakan pada penelitian ini ingin mengetahui topik yang dibahas jika authornya adalah partai dan jika authornya adalah kubu. Salah satu keterbatasan dari ATM adalah tidak mampu

memodelkan topik dengan baik pada saat author-nya kubu capres cawapres, atau jika author-nya hanya berjumlah sedikit. Sehingga pada penelitian ini, untuk mengetahui topik yang dibahas berdasarkan partai menggunakan metode ATM, dan untuk mengetahui topik yang dibahas berdasarkan kubu pendukung capres cawapres menggunakan Latent Dirichlet Allocation (LDA). Proses ini dilakukan dengan menggunakan *library* Gensim, yang kemudian setelah model terbentuk akan dievaluasi menggunakan nilai *perplexity*. Pada proses ini, meskipun terdapat perbedaan metode yang digunakan antara author partai dan kubu pendukung, tapi secara garis besar cara yang dilakukan hampir sama, yaitu eksperimen untuk mengubah nilai parameter pada setiap kali dilakukan pembentukan model. Dari setiap parameter yang dimasukkan selama pembentukan model, akan dilakukan evaluasi nilai *perplexity* menggunakan modul *logging*, dan model dikatakan baik apabila nilai *perplexity* rendah dan stabil. Proses topic modeling dilakukan pada Gambar 4.7.



**Gambar 4.7** Proses Permodelan Topik dengan ATM dan LDA

#### 4.6 Validasi Permodelan Topik

Validasi model dilakukan untuk memastikan model yang sudah dibuat dari topic modeling pada dokumen, dan kata-kata yang ada pada suatu topik sudah sesuai. Beberapa hal yang perlu dianalisis untuk memvalidasi model topik antara lain:

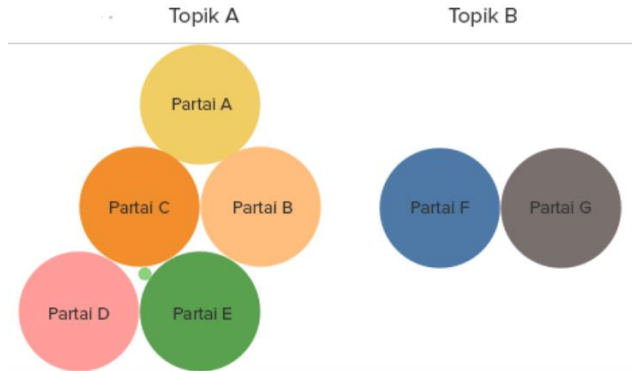
1. Jumlah iterasi dan pass sesuai untuk membentuk model topik.
2. Jumlah topik yang sesuai dengan nilai rata-rata dan standar deviasi *perplexity* terkecil.
3. Kata-kata yang mendukung setiap topik apakah sudah saling berkaitan dan cukup merepresentasikan suatu topik.

#### **4.7 Analisis Topik**

Proses selanjutnya adalah menganalisis topik yang didapatkan setelah dilakukan proses author-topic model maupun dengan LDA. Proses analisis ini dilakukan dengan cara menerjemahkan setiap kata yang membentuk suatu topik beserta probabilitasnya. Topik yang dianalisis adalah topik yang memiliki rata-rata dan standar deviasi yang lebih stabil dibandingkan dengan topik-topik lainnya. Topik tersebut dianalisis sesuai dengan pengetahuan umum mengenai isu-isu yang kemungkinan dibahas di media sosial khususnya Twitter selama masa Pemilu 2019. Pembuatan kategori topik disesuaikan dengan setiap kata yang mendukung topik tersebut, dan disesuaikan setiap hubungan dan keterkaitan antar kata dalam 1 topik yang sama.

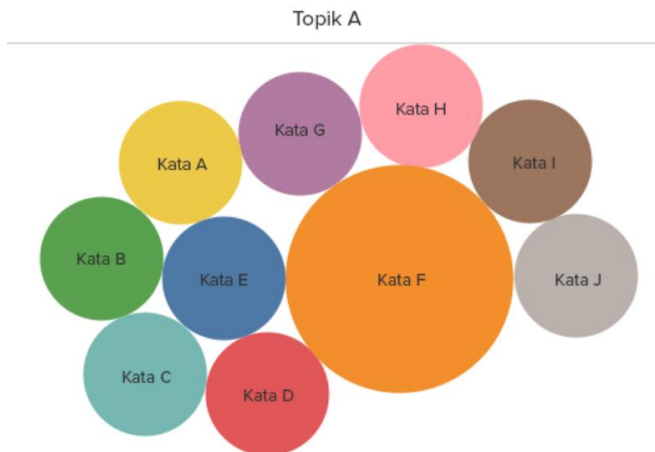
#### **4.8 Desain Visualisasi Permodelan Topik**

Untuk desain visualisasi hasil permodelan topik yang sudah dilakukan, akan dibuat beberapa grafik dalam bentuk *packed bubbles* atau beberapa lingkaran yang saling berdekatan dengan besar ukuran yang berbeda-beda. Ukuran lingkaran ditentukan oleh probabilitas dari masing-masing *item* yang berkaitan dengan visualisasi tersebut.



**Gambar 4.8 Desain Visualisasi Author pada tiap Topik**

Gambar 4.8 menunjukkan desain yang akan dibuat untuk memvisualisasikan *author* partai yang memiliki probabilitas dalam membahas suatu topik. Dalam desain visualisasi tersebut setiap partai ditampilkan sebagai 1 lingkaran, yang besarnya lingkaran tersebut tergantung pada besar probabilitas *author* dalam topik tersebut.



**Gambar 4.9 Desain Visualisasi Detail Kata tiap Topik**

Lalu untuk desain visualisasi yang menampilkan detail kata pada tiap topik akan dibuat seperti Gambar 4.9. Setiap kata yang

menyusun Topik A akan ditampilkan dalam bentuk lingkaran, dimana besar ukuran lingkaran akan menyesuaikan dengan tingkat probabilitas dari masing-masing kata yang menyusun topik tersebut. Semakin besar probabilitas kata tersebut dalam suatu topik, maka semakin besar juga ukuran dari lingkaran tersebut.



## BAB V IMPLEMENTASI

Bab ini menjelaskan hasil implementasi dari perancangan yang sudah dilakukan sebelumnya. Hasil yang dijelaskan pada bab ini adalah hasil percobaan terhadap data berupa *tweet* yang sudah dikumpulkan sebagai acuan dalam penelitian.

### 5.1 Lingkungan Implementasi

Pada pembuatan pengelompokan dan visualisasi topik tweet dari akun caleg selama Pemilu 2019, diperlukan perangkat yang digunakan untuk menunjang jalannya penelitian. Daftar perangkat yang digunakan ditunjukkan pada Tabel 5.1 dan Tabel 5.2

**Tabel 5.1 Spesifikasi Perangkat Keras**

No	Hardware	Spesifikasi
1.	Jenis Laptop	ACER E5-476G
2.	Processor	Intel(R) Core(TM) i5-7200U
3.	RAM	4 GB
4.	Hard disk	1 TB

**Tabel 5.2 Spesifikasi Perangkat Lunak dan Library**

<b>Bahasa Pemrograman</b>	<ul style="list-style-type: none"><li>• Python 3.7</li></ul>
<b>Text Editor</b>	<ul style="list-style-type: none"><li>• Visual Studio Code</li><li>• Sublime Text Editor</li></ul>
<b>Software / Tools</b>	<ul style="list-style-type: none"><li>• Microsoft Office 365</li><li>• Tableau 2018.3</li><li>• MySQL DBMS</li></ul>
<b>Library</b>	<ul style="list-style-type: none"><li>• NLTK 3.3.0</li><li>• Sastrawi 1.0.1</li></ul>

	<ul style="list-style-type: none"> <li>• Gensim 3.7.1</li> <li>• Pandas 0.23.4</li> <li>• Tweepy 3.7.0</li> <li>• Mysql-connector 2.1.6</li> </ul>
--	--

## 5.2 Pengambilan dan Penyimpanan Data pada Basis Data

Pengambilan data adalah tahap pertama yang dilakukan untuk mengumpulkan *tweet* dari akun Twitter caleg yang sudah dikumpulkan sebelumnya. Jumlah akun Twitter yang dikumpulkan pada setiap partai pendukung kedua kubu adalah 15 akun. Setiap akun yang dikumpulkan sebelumnya telah melewati tahap verifikasi untuk memastikan akun tersebut layak dikumpulkan *tweetnya* dan dapat diolah pada tahapan selanjutnya.

### 5.2.1 Pengumpulan Data Akun

Dalam penelitian ini setiap partai dan kubu pendukung capres dan cawapres diberikan kode atau id baru yang digunakan untuk mempermudah proses identifikasi dan pengolahan data selama melakukan permodelan topik. Kode partai yang berawalan 1 menandakan partai tersebut mendukung atau mengusung kubu Jokowi-Amin, dan kode partai yang berawalan 2 mendukung pasangan Prabowo-Sandi. Tabel 5.3 dan Tabel 5.4 menunjukkan daftar kode partai beserta nama partai pada setiap kubu pendukung capres dan cawapres pada Pemilu 2019.

**Tabel 5.3 Kode Kubu Capres-Cawapres**

<b>Kode Kubu</b>	<b>Nama Kubu Pendukung</b>
01	Joko Widodo – Ma'ruf Amin
02	Prabowo Subianto – Sandiaga Uno

Tabel 5.4 Detail Kode Partai

<b>Kode Partai</b>	<b>Nama Partai</b>
101	Partai Kebangkitan Bangsa (PKB)
102	Partai Demokrasi Indonesia Perjuangan (PDIP)
103	Partai Golongan Karya (Golkar)
104	Partai Nasdem
105	Partai Persatuan Pembangunan (PPP)
106	Partai Solidaritas Indonesia (PSI)
107	Partai Persatuan Indonesia (Perindo)
108	Partai Hati Nurani Rakyat (Hanura)
109	Partai Keadilan dan Persatuan Indonesia (PKPI)
201	Partai Gerakan Indonesia Raya (Gerindra)
202	Partai Demokrat
203	Partai Amanat Nasional (PAN)
204	Partai Keadilan Sejahtera (PKS)
205	Partai Berkarya

Akun Twitter caleg dikumpulkan secara manual dengan menggunakan data yang ada pada website KPU. Setiap nama caleg dicari secara manual, kemudian dibandingkan dengan data dan foto setiap profil caleg yang ada pada website KPU, kemudian dilihat untuk memastikan bahwa akun tersebut masih aktif digunakan dan layak dijadikan bahan dalam penelitian. Selanjutnya setelah dilakukan validasi, didapatkan hasil rekapitulasi akun Twitter caleg pada Tabel 5.5

**Tabel 5.5 Rekapitulasi Validasi Akun Twitter**

Jumlah partai	14
Total akun Twitter yang dikumpulkan	210
Akun valid	208
Akun tidak valid	2

### **5.2.2 Crawling Data**

Dalam proses pengambilan data menggunakan library Tweepy yang mengakses API dari Twitter, dan kemudian langsung disimpan pada database. Kode 5.1 menunjukkan proses crawling data dari Twitter yang dijelaskan pada subbab 4.2.

```

1. def getPage(username):
2.     startDate = dt.datetime(2018, 5, 22, 0, 0,
3.     0)
4.     endDate = dt.datetime(2019, 5, 22, 0, 0, 0)
5.     )
6.     try:
7.         print(username + " mulai")
8.         for page in tweepy.Cursor(api.user_timeline, screen_name=username, tweet_mode="extended").pages():
9.             for t in page:
10.                if t.created_at < endDate and t.created_at > startDate:
11.                    username = t.user.screen_name
12.                    id_tweet = t.id
13.                    tgl_tweet = t.created_at
14.                    text = t.full_text
15.                    sql_insert_query = """ INSERT INTO `tweet_baru`
16.                    (`id_tweet`, `username`, `text`, `tgl_tweet`) VALUES (%s,%s,%s,%s)"""
17.                    insert_tuple = id_tweet, username, text, tgl_tweet
18.                    cursor = connection.cursor()
19.                    result = cursor.execute(sql_insert_query, insert_tuple)
20.                    connection.commit()
21.                    # print ("Record inserted successfully into tweet_baru table")
22.                    print(username + " selesai")
23.                except tweepy.TweepError:
24.                    print(username + " Failed to run the command on that user, Skipping...")

```

### Kode 5.1 Kode untuk melakukan Crawling Tweet

Dalam melakukan crawling data, daftar setiap username akun Twitter yang sudah dikumpulkan dimasukkan kedalam sebuah file csv. Tweet yang diambil adalah tweet yang dibuat pada tanggal 22 Mei 2018 sampai dengan 22 Mei 2019, yang

ditunjukkan pada baris 2-3. Selanjutnya hasil crawling dari setiap tweet akan otomatis masuk ke dalam database yang sudah dibuat sebelumnya. Beberapa atribut yang masuk ke dalam database antara lain id tweet, username, tweet, dan tanggal tweet tersebut dibuat, yang ditunjukkan pada baris 10-13.

```

1. def getTweets():
2.     df = pd.read_csv("username.csv", sep=";",
3.                     error_bad_lines=False, encoding='cp1252',
4.                     warn_bad_lines=False)
5.     usernames = df.username
6.     for user in usernames:
7.         utils.getPage(user)

```

**Kode 5.2 Fungsi untuk melakukan Crawling**

### 5.3 Data Pre-Processing

Setelah semua tweet dikumpulkan dalam database, sebelum data diolah lebih lanjut, data perlu dilakukan pre-processing, yang telah dijelaskan pada subbab 4.4. Berikut adalah contoh kode saat pre-processing *tweet*:

1. Case folding atau melakukan pengubahan format huruf menjadi *lowercase* atau bentuk huruf kecil. Kode 5.3 menunjukkan proses perubahan bentuk huruf menjadi *lowercase*.

```

1. lowerdocs = str(i.lower())

```

**Kode 5.3 Kode untuk Lowercase**

2. Penghapusan beberapa karakter khusus seperti emoji, hashtag, dan link. Kode 5.4 menunjukkan proses penghapusan karakter-karakter yang tidak diperlukan.

```

1. hapuslink = re.sub(r'\w+:\{\2}[\d\w-
]+(\.[\d\w-
]+)*(?:\:(?:\/[^\s/]*))*', '', lowerdocs)
2. hapushashtag = re.sub(r"#(\w+)", '', hapuslink
)
3. hapusline = re.sub("\n|\r", '', hapushashtag)
4. hapususername = re.sub('@[^\s]+', '', hapusline)

5. hapusemoji = give_emoji_free_text(hapususername)
6. stemmeddocs = stemmer.stem(hapusemoji)
7. hapussimbol = re.sub('[!@$,#&_*~.-
:?"();%/]', ' ', stemmeddocs)

```

#### Kode 5.4 Kode untuk Penghapusan Karakter, Hashtag, Link, dan Emoji

- Perubahan kata menjadi bentuk kata dasar dengan menghilangkan imbuhan pada kata atau *stemming*. Stemming dilakukan dengan bantuan library Sastrawi, yaitu dengan method StemmerFactory(). Kode 5.5 menunjukkan bagaimana *stemming* dilakukan.

```

1. from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
2.
3. factory = StemmerFactory()
4. stemmer = factory.create_stemmer()

```

#### Kode 5.5 Kode untuk Stemming

- Pemisahan kalimat dalam 1 *tweet* menjadi bentuk per kata atau tokenization. Tokenization dilakukan menggunakan bantuan library NLTK yaitu menggunakan RegexpTokenizer. Kode 5.6 menunjukkan bagaimana tokenization dilakukan.

```

1. from nltk.tokenize import RegexpTokenizer
2. tokenlist_data = []
3. tokenizer = RegexpTokenizer('\w+|\$[\d\.]+|\S+'
4.
5.         tokens = tokenizer.tokenize(hapussimbo
6.         tokenWithoutInt = [item for item in to
           kens if not item.isdigit()])

```

#### Kode 5.6 Kode untuk Tokenization

5. Penghapusan kata-kata yang sering muncul tapi tidak memiliki makna yang signifikan atau *stopwords removal*. Penghapusan ini dilakukan dengan mengurutkan kata berdasarkan jumlah kemunculan, dan dari urutan tersebut dilihat apakah kata-kata yang sering muncul memiliki makna yang penting atau tidak. Daftar *stopwords* yang dihapus pada tugas akhir ini dapat dilihat di LAMPIRAN C. List Stopwords. Kode 5.7 menunjukkan bagaimana *stopwords removal* dilakukan. Pada mulanya dilakukan inisiasi untuk mendaftarkan apa saja kata yang termasuk dalam kategori stopword, kemudian kata-kata tersebut dihapus dari dokumen.

```

1. list_stopword = []
2. with open('stopwords.txt') as stopword:
3.     for line in stopword:
4.         list_stopword.append(line.strip())

```

#### Kode 5.7 Kode untuk menghapus Stopword



Berikut adalah kode yang digunakan pada pre-processing secara keseluruhan.

```
1. def preproses(df):
2.     docs = df[3]
3.     doc_ids = df[0].astype(str).replace('\.0',
4.     '', regex=True)
5.     doc_ids = doc_ids.apply(str)
6.     list_stopword = []
7.     with open('stopwords.txt') as stopword:
8.         for line in stopword:
9.             list_stopword.append(line.strip())
10.
11.     factory = StemmerFactory()
12.     stemmer = factory.create_stemmer()
13.
14.     tokenlist_data = []
15.     tokenizer = RegexpTokenizer('\w+|\$[\d\.]+
16.     |\S+')
17.     print('Praproses mulai')
18.     print('Jumlah dokumen sebelum praproses %d
19.     ' % len(docs))
```

```

19.     for i in docs:
20.         lowerdocs = str(i.lower())
21.         hapuslink = re.sub(r'\w+:\/\/{2}[\d\w-
    ]+(\.[\d\w-
    ]+)*(?:\:\/\/[^\s/]*))*', '', lowerdocs)
22.         hapushashtag = re.sub(r"#(\w+)", '', h
    apuslink)
23.         hapusline = re.sub("\n|\r", '', hapushas
    htag)
24.         hapususername = re.sub('@[^\s]+', '', ha
    pusline)
25.         hapusemoji = give_emoji_free_text(hapu
    susername)
26.         stemmeddocs = stemmer.stem(hapusemoji)
27.         # stemming process
28.         hapussymbol = re.sub('[!@$,#&_*.~.-
    :?"]();%/]', ' ', stemmeddocs)
29.         # print(hapusemoji)
30.         tokens = tokenizer.tokenize(hapussimbo
    l)
31.         tokenWithoutInt = [item for item in to
    kens if not item.isdigit()]
32.         stopped_tokens = [i for i in tokenWith
    outInt if not i in list_stopword]
33.         hapusstring = [x for x in stopped_toke
    ns if x]
34.         tokenlist_data.append(hapusstring)
35.         # print(hapusstring)
36.         print('Jumlah dokumen sebelum stopwords %d
    ' % len(tokens))
37.         print('Jumlah dokumen setelah praproses %d
    ' % len(tokenlist_data))
38.         print("praproses selesai")

```

#### Kode 5.8 Pre-processing secara keseluruhan

Kemudian potongan Kode 5.9 digunakan untuk memanggil fungsi preproses yang sudah dibuat sebelumnya. Tweet diambil langsung dari database, sehingga memerlukan penghubung antara python dengan database yang ada pada MySQL.

```

1. def ambilTweet():
2.     try:
3.         sql_select_Query = "select * from twee
         t_baru"
4.         cursor = connection .cursor()
5.         cursor.execute(sql_select_Query)
6.         df = DataFrame(cursor.fetchall())
7.
8.         print(df.columns)
9.         utils.preproses(df)
10.
11.     except Error as e :
12.         print ("Error while connecting to MySQL
L", e)
13.     finally:
14.         #closing database connection.
15.         if(connection .is_connected()):
16.             connection.close()
17.             print("MySQL connection is closed"
)

```

**Kode 5.9 Fungsi untuk melakukan Pre-processing**

## 5.4 Pemrosesan Data

Dalam penelitian ini dijelaskan pemrosesan data dengan menggunakan LDA dan ATM. Secara garis besar, proses yang dilakukan hampir sama karena ATM sendiri adalah pengembangan dari LDA. Berikut akan dijelaskan beberapa tahapan yang dilakukan selama melakukan pemrosesan data, yang telah disebutkan pada subbab 4.5.

### 5.4.1 Proses Memuat Data

Pada tahapan pemrosesan data, *tweet* yang sebelumnya telah dilakukan pre-processing dikumpulkan dan dimuat untuk selanjutnya dilakukan pemrosesan data. Dokumen *tweet* yang sudah dilakukan pre-processing dimuat dengan menggunakan library csv, yang ditunjukkan pada baris 4-11. Kode 5.10 menunjukkan proses load data *tweet* hingga siap diolah.

```

1. docs=[]
2. i=0
3. tokenizer = RegexpTokenizer(r'\w+')
4. with open('baru-kubu' + kubu + '-
   praproses.csv') as raw:
5.     reader = csv.reader(raw)
6.     for row in reader:
7.         gabung = ''.join(row[0])
8.         token = tokenizer.tokenize(gabung)
9.         if len(token)>2:
10.            docs.append(token)
11.            i=i+1

```

**Kode 5.10 Kode untuk memuat tweet**

Secara umum, *tweet* yang akan diolah adalah *tweet* yang pada 1 dokumen terdapat lebih dari 2 kata, yang ditunjukkan pada baris 9-11. Jika ada *tweet* yang tidak memenuhi syarat, maka akan secara otomatis tidak diolah dalam pembentukan model topik. Hal ini dilakukan untuk membuat bentuk corpus menjadi lebih padat sehingga membantu proses pembentukan model topik kedepannya.

### 5.4.2 Proses Pembuatan Dictionary dan Corpus

Pada subbab 4.5.1 telah dijelaskan mengenai pembentukan *dictionary* dan *corpus*. Setelah data berhasil dimuat, selanjutnya perlu dibuat sebuah *dictionary*. Dictionary bertujuan untuk mempermudah dalam mengatur kata apa saja yang dapat digunakan selama pembentukan model berlangsung. Dengan bentuk *dictionary*, setiap kata memiliki sebuah index yang unik yang mempermudah saat dilakukan pembentukan model topik. Pembuatan *dictionary* ditunjukkan pada Kode 5.11.

```

1. dictionary = Dictionary(docs)
2. max_freq = 0.9
3. min_wordcount = 20
4. dictionary.filter_extremes(no_below=min_wordcount, no_above=max_freq)
5. _ = dictionary[0]

```

#### Kode 5.11 Kode untuk membuat Dictionary

Pada baris 2-3 menunjukkan pembentukan *dictionary* dengan frekuensi kata maksimal sering muncul adalah 90% dari total keseluruhan dokumen, dan jumlah minimal kata tersebut muncul adalah 20 kali. Penentuan nilai `max_freq` dan `min_wordcount` disesuaikan dengan bentuk dan jumlah data pada masing-masing penelitian. Selain itu, penentuan jumlah ini juga ditentukan oleh contoh hasil saat pembentukan model pertama kali dilakukan. Pada saat model pertama kali dibentuk, kemungkinan akan terbentuk model dengan kata-kata yang tidak terlalu relevan adalah sangat mungkin, sehingga perlu dilakukan percobaan untuk menentukan parameter yang pas.

Setelah *dictionary* terbentuk, selanjutnya perlu dilakukan pembentukan corpus atau menjadi bentuk vector kata dari dokumen. Kode pembentukan corpus ditunjukkan pada Kode 5.12.

```

1. corpus = [dictionary.doc2bow(doc) for doc in docs]

```

#### Kode 5.12 Kode Pembentukan Corpus

Setelah corpus terbentuk, tahapan selanjutnya adalah pembentukan model topik. Pembentukan model topik dilakukan dengan 2 metode, yaitu untuk author partai menggunakan ATM dan untuk author kubu pendukung capres cawapres menggunakan LDA.

### 5.4.3 Pencatatan Nilai Perplexity

Setiap perubahan nilai perplexity yang didapatkan dari setiap percobaan jumlah topik kemudian dicatat menggunakan

*logging*, yang kemudian menjadi bahan analisis berapa jumlah topik yang tepat. Kode 5.13 digunakan untuk mencatat setiap perubahan perplexity pada setiap penentuan jumlah topik pada setiap passesnya.

```

1. import logging
2. logger = logging.getLogger()
3. fhandler = logging.FileHandler(filename='lda+k
   ubu'+ kubu + '-logging-
   ' + str(jmltopik) + '.csv', mode='w')
4. formatter = logging.Formatter('%(asctime)s - %
   (name)s - %(levelname)s - %(message)s')
5. fhandler.setFormatter(formatter)
6. logger.addHandler(fhandler)
7. logger.setLevel(logging.DEBUG)

```

**Kode 5.13 Kode Pencatatan *Perplexity***

#### **5.4.4 Pembentukan Topik dengan Author Kubu Pendukung Capres dan Cawapres**

Pembentukan model topik dengan author berdasarkan kubu pendukung capres dan cawapres dilakukan dengan menggunakan Latent Dirichlet Allocation (LDA). Pembentukan model ini terkait dengan penjelasan pada subbab 4.5.3. Dengan jumlah author yang hanya berasal dari 2 kubu, cara pembentukan model topik dengan LDA dilakukan dengan cara memisahkan dokumen atau *tweet* yang dibuat dari kubu 01 dan kubu 02. Tweet dari kubu 01 dikumpulkan dari partai yang mendukung dan mengusung paslon 01, begitu juga berlaku sebaliknya pada kubu 02, dan masing-masing dikumpulkan pada sebuah file csv yang kemudian digunakan untuk pembentukan model topik. Proses pembuatan model topik dilakukan dua kali, yaitu sesuai dengan jumlah kubu pendukung capres dan cawapres. Kode yang digunakan untuk pembentukan model topik menggunakan LDA adalah sebagai berikut.

```
1. lda1 = LdaModel(corpus=corpus,  
2.                 num_topics=jmltopik,  
3.                 id2word=dictionary.id2token,  
4.                 alpha = 'auto',  
5.                 eta='auto',  
6.                 passes=1000,  
7.                 eval_every=10,  
8.                 iterations=10000)
```

#### Kode 5.14 Kode Permodelan Topik dengan LDA

Beberapa hal yang perlu diperhatikan selama pembentukan model topik adalah parameter yang berkaitan dengan jumlah topik, *passes*, dan iterasi. Pada percobaan untuk mengetahui jumlah topik yang terbentuk, dilakukan perubahan parameter topik mulai dari 2 sampai dengan 7 topik. Penentuan jumlah *passes* sebanyak 1000 kali dilakukan untuk membuat model topik yang lebih stabil.

#### 5.4.5 Pembentukan Topik dengan Author Partai

Pembentukan model topik dengan author 14 partai dilakukan menggunakan Author-Topic Model. Pembentukan model ini terkait dengan penjelasan pada subbab 4.5.3. Terdapat beberapa perbedaan dengan pembentukan model topik dengan LDA, salah satunya perlu dibentuk **author2doc** atau dilakukan mapping index dokumen yang menandakan author dari masing-masing *tweet*, yaitu berdasarkan 14 partai yang mengikuti Pemilu 2019. Kode 5.15 menunjukkan proses pembuatan model topik dengan menggunakan ATM yang diawali dengan proses memuat data *tweet* dan juga data partai sebagai author.

```

1. docs=[]
2. doc_ids=[]
3. i=0
4. tokenizer = RegexpTokenizer(r'\w+')
5. with open('baru-praproses4-
   stopwords.csv') as raw:
6.     reader = csv.reader(raw)
7.     for row in reader:
8.         gabung = ''.join(row[1])
9.         token = tokenizer.tokenize(gabung)
10.        doc_id = row[0]
11.        if len(token)>4:
12.            docs.append(token)
13.            doc_ids.append(doc_id)
14.            i=i+1
15.
16. author2doc = dict()
17. i = 0
18. with open('partai.csv', errors='ignore', encoding="utf8") as author:
19.     readCSV = csv.reader(author, delimiter=';'
   )
20.     next(readCSV, None)
21.     for contents in readCSV:
22.         authorname = contents[1]
23.         authorname = re.sub('\s', ' ', authorname)
24.         ids = contents[0]
25.         if not author2doc.get(authorname):
26.             author2doc[authorname] = []
27.             author2doc[authorname].append(ids)
28.
29. doc_id_dict = dict(zip(doc_ids, range(len(doc_ids))))
30.
31. for a, a_doc_ids in author2doc.items():
32.     for i, doc_id in enumerate(a_doc_ids):
33.         author2doc[a][i] = doc_id_dict[doc_id]

```

#### Kode 5.15 Kode Pembuatan Author-to-doc

Setelah *tweet* dan *partai* berhasil dimuat dan telah dipetakan dokumen berdasarkan *authorname*, selanjutnya dilakukan proses pembentukan dictionary dan corpus. Pembentukan



dictionary dan corpus dilakukan dengan cara yang sama yang ada pada Kode 5.11 dan Kode 5.12. Setelah dictionary dan corpus terbentuk, selanjutnya dibentuk model topik menggunakan Author-Topic Model. Kode 5.16 menunjukkan pembuatan model topik dengan ATM.

```
1. atm1 = AuthorTopicModel(corpus=corpus,
2.                          num_topics=jmltopik,
3.                          id2word=dictionary.id2token,
4.                          author2doc=author2doc,
5.                          alpha = 'auto',
6.                          passes=1000,
7.                          eval_every=10,
8.                          iterations=10000)
```

#### Kode 5.16 Kode Permodelan Topik dengan ATM

Selama melakukan pembentukan model topik, jumlah topik menjadi salah satu parameter yang dilakukan perubahan nilainya. Pada penelitian ini, dilakukan percobaan mulai dari 2 hingga 7 topik untuk mengetahui jumlah topik yang dapat diketahui dari *tweet* yang dibuat oleh 14 partai tersebut. Jumlah topik dapat diubah dengan mengubah nilai parameter pada `num_topics` pada baris ke-2.

#### 5.4.6 Penyimpanan Model Topik

Setelah setiap kali selesai dilakukan percobaan pembentukan model topik, dilakukan penyimpanan model untuk mempermudah pengolahan model topik kedepannya. Kode 5.17 dan Kode 5.18 menunjukkan proses penyimpanan model baik model LDA maupun ATM.

```
1. lda1.save('lda+kubu' + kubu + '-'
            + str(jmltopik) + 'topik.ldamodel')
```

#### Kode 5.17 Kode Penyimpanan Model Topik LDA

```
1. atm1.save('atm+' + str(jmltopik) + 'topik-
            baru3.atmodel')
```

#### Kode 5.18 Kode Penyimpanan Model Topik ATM

## 5.5 Validasi Permodelan Topik

Validasi setiap model topik yang sudah dibuat dilakukan dengan melakukan perhitungan pada perplexity yang tercatat pada setiap passes yang dilakukan, yang sudah dijelaskan pada subbab 4.6. Model yang dipilih adalah model dengan nilai rata-rata dan standar deviasi paling stabil dibandingkan dengan model lain yang sudah dibuat. Nilai perplexity diambil dari pencatatan yang dilakukan menggunakan logging, dan kemudian diolah menggunakan bantuan Microsoft Excel untuk mengetahui rata-rata dan standar deviasi setiap model yang telah dibuat, yang ditunjukkan pada Kode 5.19 baris 5-7. Kode 5.19 secara keseluruhan digunakan untuk mengambil nilai perplexity pada tiap passes kemudian disimpan dalam bentuk file csv.

```

1. lines = [line.strip() for line in open('atm+lo
   logging-' + str(jmltopik) + '-baru3.csv')]
2.
3. for x in lines:
4.     try:
5.         found = re.search('per-
   word bound, (.+?) perplexity', x).group(1)
6.         f = open('atm+' + str(jmltopik) + 'per
   plexity-baru3.csv', 'a')
7.         f.write("\n"+found)
8.     except AttributeError:
9.         found = 'tidak ada'
10. print('perplexity udah')
```

**Kode 5.19** Kode untuk mengambil Nilai *Perplexity*

## 5.6 Analisis Topik

Analisis topik dilakukan untuk menentukan berapa jumlah topik yang dipilih. Implementasi analisis topik dilakukan berdasarkan perancangam pada subbab 4.7. Setelah melakukan perhitungan rata-rata dan standar deviasi perplexity, analisis pertama yang perlu dilakukan adalah analisis mengenai kata-kata dan juga probabilitas kata tersebut dalam suatu topik. Proses analisis dilakukan dengan menyimpan setiap hasil kata dan probabilitas authornya kedalam beberapa file, yang kemudian dapat

dianalisis persebaran kata dan probabilitasnya, yang ditunjukkan pada baris 7-8 pada Kode 5.20.

```

1. with open('atm+' + str(jmltopik) + 'topik-
   detail-baru3.csv','w') as tulisFile:
2.     tulisFileWriter = csv.writer(tulisFile
   ,lineterminator='\n')
3.     d=0
4.     for items in top_topics:
5.         v=0
6.         for items in top_topics[0][0]:
7.             h = top_topics[d][0][v][0]
8.             w = top_topics[d][0][v][1]
9.             v=v+1
10.            tulisFileWriter.writerow([d,h,
   w])
11.            if v ==10:
12.                break
13.            d=d+1
14.    tulisFile.close()
15.    print('detail topik udah')
```

#### Kode 5.20 Kode Penyimpanan Detail Kata-Kata pada Topik

Setelah melakukan analisis pada kata-kata penyusun suatu topik, selanjutnya perlu dilakukan analisis mengenai probabilitas author pada suatu topik. Hal ini dilakukan untuk mengetahui dari partai mana saja yang banyak membicarakan topik tersebut selama Pemilu 2019. Kode 5.21 digunakan untuk menyimpan probabilitas suatu partai dalam suatu topik. Baris 6-7 digunakan untuk menyimpan *author* beserta probabilitasnya dalam suatu partai.

```

1. with open('atm+' + str(jmltopik) + 'probabilit
   as-baru3.csv','w') as tulisFile:
2.     tulisFileWriter = csv.writer(tulisFile,lin
   eterminator='\n')
3.     initopik = 0
4.     i=0
5.     for keys in author2doc:
6.         cobrauthor = author_vecs[i][0][0]
7.         cobatopik  = author_vecs[i][0][1]
8.         if len(author_vecs[i])>1:
9.             d=0
10.            for items in author_vecs[i]:
11.                cobrauthor2 = author_vecs[i][d
12.                ][0]
13.                cobatopik2  = author_vecs[i][d
14.                ][1]
15.                tulisFileWriter.writerow([keys
16.                ,cobatopik2,cobrauthor2])
17.                d=d+1
18.            else:
19.                tulisFileWriter.writerow([keys,cob
20.                atopik,cobrauthor])
21.            i=i+1
22.        tulisFile.close()

```

**Kode 5.21 Kode Penyimpanan Author pada Topik**

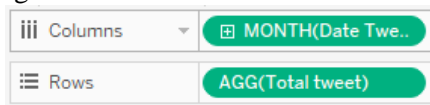
## 5.7 Visualisasi Permodelan Topik

Setelah didapatkan model dengan jumlah topik yang terbaik, selanjutnya dilakukan proses visualisasi setiap topik yang sudah dibuat. Setiap hasil topik yang didapatkan kemudian disimpan dalam 1 basis data, yang kemudian dilakukan pembuatan visualisasi dengan Tableau. Visualisasi pada Tableau dilakukan dengan membagi pada beberapa workbook. Hal ini dilakukan untuk mempercepat proses dalam mengolah visualisasi dan juga keterbatasan Tableau Public yang hanya bisa menyimpan data tidak lebih dari 15.000.000 row. Beberapa workbook yang dibuat pada visualisasi antara lain:

1. Workbook rekap *tweet* Pemilu 2019 berisi informasi umum mengenai rekap dan hasil akhir *tweet* yang berhasil

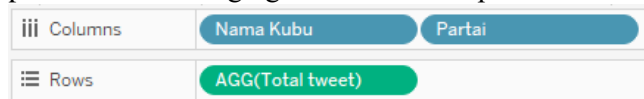
dikumpulkan. Dalam visualisasi ini dapat diketahui informasi total *tweet* yang berhasil dikumpulkan, waktu dimana akun Twitter caleg sering membagikan *tweet*, dan informasi lainnya. Pembuatan workbook dilakukan dengan menggabungkan 3 tabel pada database caleg, antara lain tabel ‘tweet\_baru’, ‘kubu’, dan ‘partai’. Measure yang dibuat pada workbook ini adalah perhitungan pada jumlah ‘text’ tweet atau ‘Total Tweet’ yang menggunakan COUNT([Text]). Dalam workbook ini memuat beberapa sheet yang digunakan untuk mengolah data, antara lain:

1. Sheet ‘Statistik Tweet’ memuat informasi mengenai statistic *tweet* pada tiap bulan selama jangka waktu yang sudah ditentukan. Pembuatan sheet dilakukan dengan menempatkan waktu pembuatan *tweet* pada kolom yaitu dalam bentuk bulan, dan agregat dari ‘Total Tweet’ pada baris atau *rows*. Kemudian untuk visualisasinya dipilih dengan menggunakan bentuk *line* atau garis.



**Gambar 5.1 Detail Sheet Statistik Tweet**

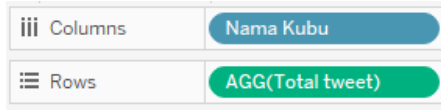
2. Sheet ‘Total Tweet per Partai’ memuat informasi jumlah *tweet* berdasarkan partai dari kedua kubu. Pada sheet ini digunakan atribut ‘Nama Kubu’ dan ‘Partai’ pada *column* dan agregat ‘Total Tweet’ pada *rows*.



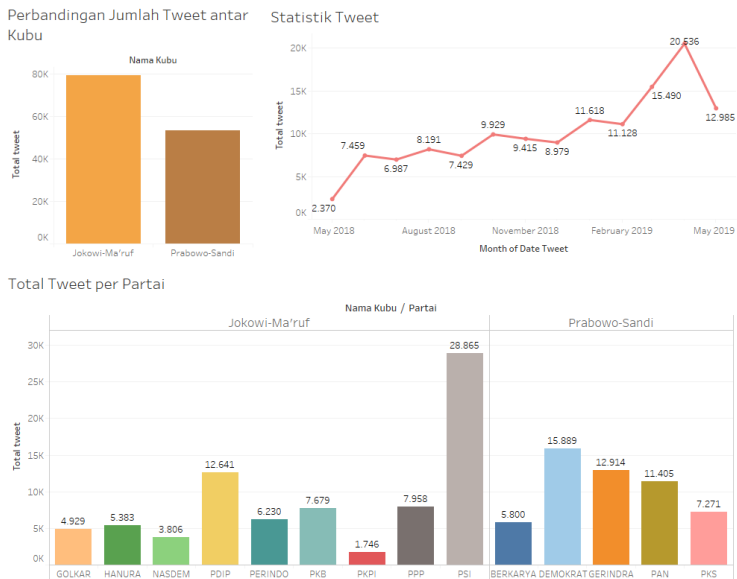
**Gambar 5.2 Detail Sheet Total Tweet per Partai**

3. Sheet ‘Perbandingan Jumlah Tweet antar Kubu’ memuat informasi perbandingan *tweet* yang sudah dibuat oleh caleg dari masing-masing kubu. Sheet ini

dibuat dengan menggunakan atribut ‘Nama Kubu’ pada *columns* dan agregat ‘Total Tweet’ pada *rows*.



**Gambar 5.3 Detail Sheet Perbandingan Jumlah Tweet per Partai**  
Setiap *sheet* berikut digabung dalam 1 dashboard dan menampilkan visualisasi data sebagai berikut:



**Gambar 5.4 Dashboard Analisis Tweet secara Umum**

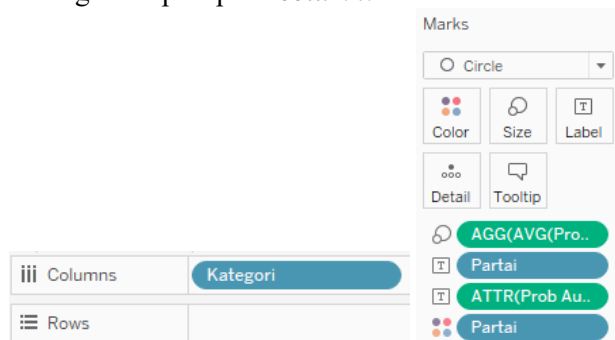
2. Workbook hasil permodelan topik, baik berdasarkan partai maupun kubu pendukung. Dalam *workbook* ini dapat diketahui topik apa saja yang banyak dibicarakan baik berdasarkan 14 partai peserta Pemilu maupun kubu capres dan cawapres, dan juga kata-kata yang mendukung topik yang dibahas oleh partai-partai atau masing-masing kubu

capres dan cawapres. Terdapat beberapa measure yang dibuat pada sheet ini antara lain:

4.  $AVG([Prob\ Author])$ , yang merupakan rata-rata dari nilai probabilitas *author* partai pada suatu topik. Digunakan untuk menampilkan probabilitas dari masing-masing partai sebagai *author* dari beberapa topik yang berhasil dibuat.
5.  $AVG([Prob\ Lda])$ , merupakan rata-rata dari nilai dari kata pada topik yang dibuat menggunakan LDA, yang menunjukkan probabilitas tiap kata yang menyusun topik tersebut.
6.  $AVG([Prob\ Topik])$ , merupakan rata-rata dari kata pada topik yang dibuat menggunakan ATM, yang memuat probabilitas tiap kata dalam menyusun suatu topik.

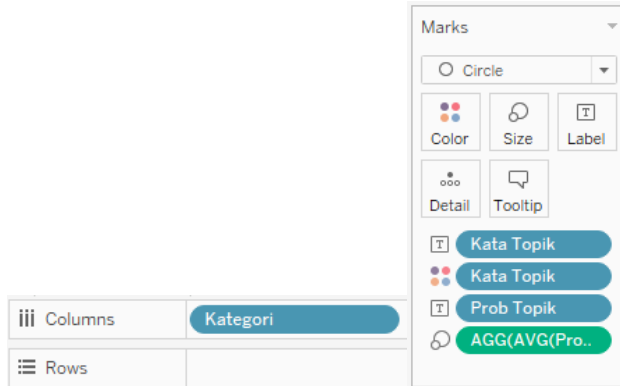
Measure tersebut kemudian digunakan untuk mengolah data pada beberapa *sheet* yang dibuat pada *workbook* ini, antara lain:

7. Sheet ‘Hasil Topik per Partai’ memuat hasil permodelan topik yang dilakukan menggunakan ATM. Sheet ini dibuat dengan menggunakan agregat ‘Prob Author’ pada bagian *marks* dan memasukkan ‘Kategori Topik’ pada *column*.



**Gambar 5.5 Detail Sheet Hasil Topik per Partai**

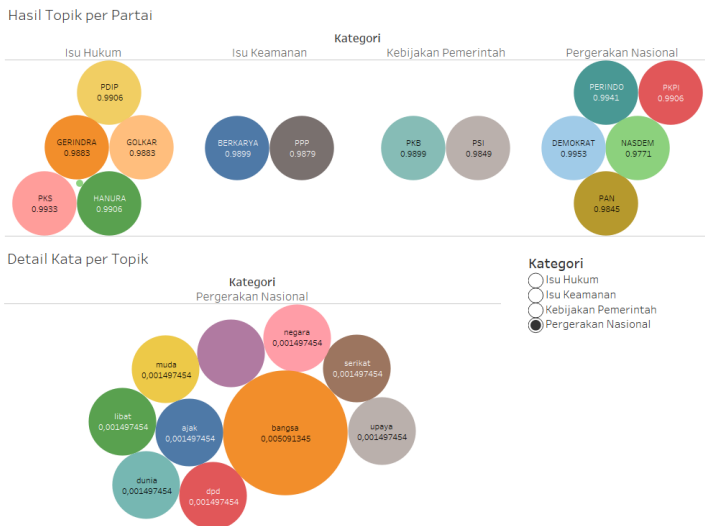
8. Sheet 'Detail Kata per Topik' memuat kata-kata yang menyusun suatu topik berdasarkan topik yang sudah terbuat. Sheet ini dibuat dengan menggunakan agregat dari 'Prob Topik' pada bagian *marks* dan memasukkan 'Kategori Topik' pada *column*.



**Gambar 5.6 Detail Sheet Detail Kata per Topik**

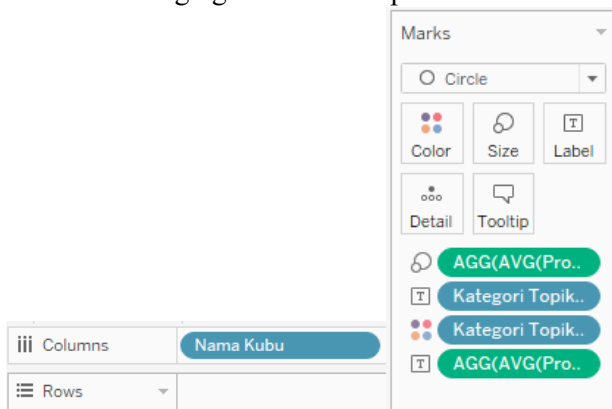
Dari kedua sheet tersebut kemudian digabung dalam 1 dashboard untuk menampilkan visualisasi 'Hasil Permodelan Topik tiap Partai' seperti pada Gambar 5.7.





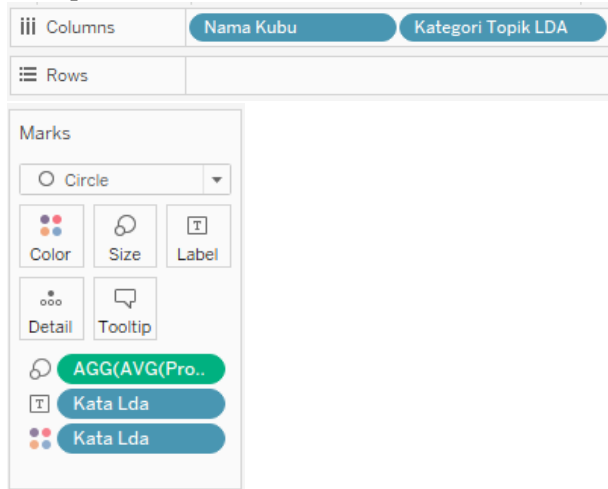
**Gambar 5.7** Dashboard Hasil Permodelan Topik tiap Partai

9. Sheet ‘Topik per Kubu’ memuat topik yang dibuat berdasarkan kubu capres dan cawapres. Sheet ini dibuat dengan memasukkan ‘Nama Kubu’ pada *columns* dan agregat ‘Prob Lda’ pada *marks*.



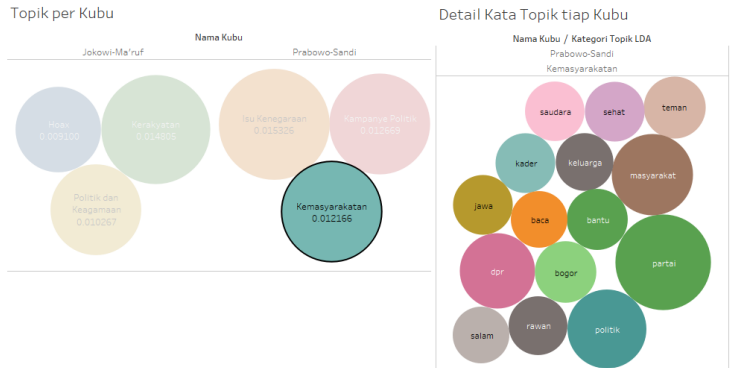
**Gambar 5.8** Detail Sheet Topik per Kubu

10. Sheet ‘Detail Kata Topik tiap Kubu’ memuat kata-kata yang ada pada setiap topik yang terbentuk berdasarkan kubu capres dan cawapres. Sheet ini dibuat dengan memasukkan ‘Nama Kubu’ dan ‘Kategori Topik Lda’ pada *columns* dan agregat ‘Prob Lda’ pada *marks*.



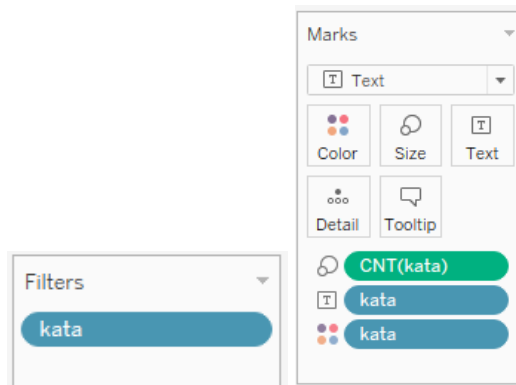
**Gambar 5.9 Detail Sheet Detail Topik tiap Kubu**

Dari kedua sheet tersebut kemudian digabung dalam 1 dashboard untuk menampilkan ‘Hasil Permodelan Topik tiap Kubu’ pada Gambar 5.10.



**Gambar 5.10 Dashboard Hasil Permodelan Topik tiap Kubu**

3. Workbook ‘Word Cloud’ digunakan untuk menampilkan word cloud atau diagram yang menunjukkan kata-kata yang paling sering disebut dalam *tweet*. Data yang digunakan pada workbook ini adalah kata-kata yang dibuat saat melakukan pembentukan *dictionary* pada proses pengolahan data. Pada workbook ini hanya terdapat 1 sheet, yang dibuat dengan memasukkan ‘Kata’ pada *marks* dan *filter* untuk menentukan berapa kata yang dimuat pada *word cloud*.



**Gambar 5.11 Detail Workbook Word Cloud**

Hasil *word cloud* yang berhasil dibuat ditunjukkan pada Gambar 5.12.



Gambar 5.12 Hasil Word Cloud

Semua hasil visualisasi dapat diakses secara *online* pada Tableau Public dengan mengakses link berikut:

<https://public.tableau.com/profile/umi.shoima#!/>

## BAB VI HASIL DAN PEMBAHASAN

Pada bab ini akan dijelaskan mengenai hasil dan pembahasan dari pembentukan model topik yang sudah dilakukan. Pembahasan dimulai dari analisis *tweet* yang sudah dikumpulkan secara umum, analisis permodelan topik yang dibuat menggunakan ATM dan LDA, hingga analisis validasi model topik yang digunakan.

### 6.1 Hasil Tweet secara Umum

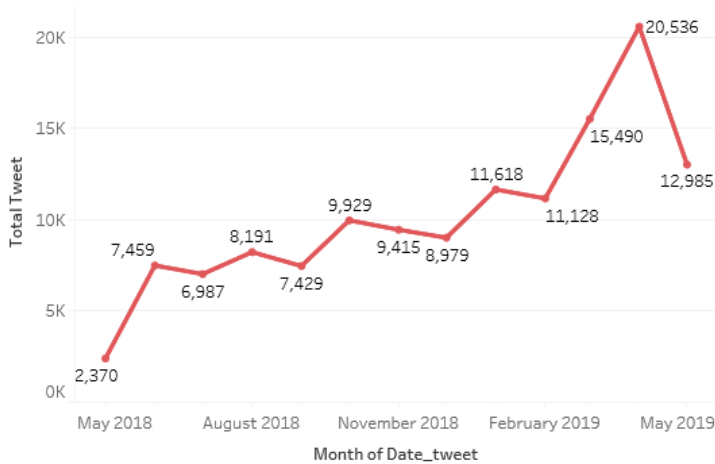
Total tweet yang didapatkan melalui *crawling* yang dilakukan sejak 22 Mei 2018 hingga 22 Mei 2019 adalah sebanyak 132.516, dengan detail *tweet* yang berhasil didapatkan tiap bulannya adalah sebagai berikut.

**Tabel 6.1 Data Total Tweet per Bulan**

<b>Tahun</b>	<b>Bulan</b>	<b>Total Tweet</b>
2018	Mei	2370
	Juni	7459
	Juli	6987
	Agustus	8191
	September	7429
	Oktober	9929
	November	9415
	Desember	8979
2019	Januari	11618
	Februari	11128
	Maret	15490
	April	20536

	Mei	12985
--	-----	-------

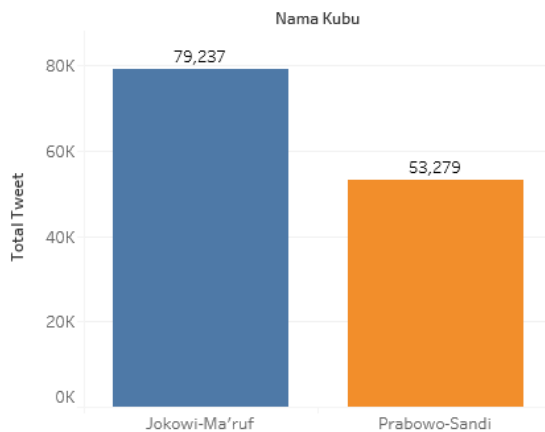
### Statistik Tweet



**Gambar 6.1 Statistik Tweet**

Dari data tersebut, dapat diketahui adanya kenaikan *tweet* yang signifikan pada bulan Februari sampai April 2019. Hal ini juga dipengaruhi oleh berlangsungnya masa kampanye yang membuat para caleg lebih aktif dalam membagikan gagasan maupun opini selama masa kampanye melalui akun Twitter mereka masing-masing. Kemudian pada bulan Mei 2019 terlihat penurunan aktivitas *tweet* yang juga ditandai dengan masa perhitungan hingga masa pengumuman pasca Pemilu 2019.

## Tweet per Kubu



**Gambar 6.2 Perbandingan Total Tweet pada tiap Kubu**

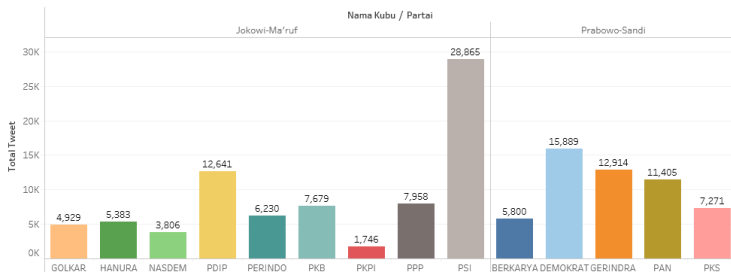
**Tabel 6.2 Hasil Tweet per Partai**

Partai	Jumlah Akun	Total Tweet
PKB	15	7679
Golkar	15	4929
Nasdem	15	3806
PDIP	15	12641
Perindo	15	6230
PKPI	15	1746
PPP	15	7958
PSI	15	28865
Berkarya	15	5800
Demokrat	15	15889
Gerindra	15	12914
PAN	15	11405

PKS	15	7271
-----	----	------

Dilihat dari perbandingan kedua kubu capres dan cawapres, dapat diketahui jika caleg dari kubu 01 atau kubu Jokowi – Ma’ruf Amin membuat lebih banyak *tweet* jika dibandingkan dengan kubu Prabowo – Sandi. Total *tweet* dari kubu 01 adalah 79.237, dan *tweet* yang dibuat oleh kubu 02 sebanyak 53.279. Hal ini tentu saja juga dipengaruhi karena jumlah partai yang bergabung dalam kubu 01 lebih banyak dibandingkan dengan kubu 02.

Total Tweet per Partai



**Gambar 6.3 Perbandingan Tweet pada tiap Partai**

Dari keseluruhan 14 partai yang menjadi peserta Pemilu 2019, Gambar 6.3 menunjukkan persebaran *tweet* yang dibagikan oleh tiap partai. Dari kubu 01 dapat diketahui bahwa caleg yang berasal dari PSI adalah yang paling banyak membagikan *tweet*, dan dari kubu 02 caleg dari Partai Demokrat adalah yang paling aktif dalam membagikan *tweet* selama Pemilu 2019.

## 6.2 Pre-processing Data

Tahapan pre-processing dilakukan dengan melakukan case folding, stemming, tokenizing, dan penghapusan stopwords. Perubahan dokumen sebelum dan setelah dilakukan pre-processing ditunjukkan pada Tabel 6.3.

**Tabel 6.3 Hasil Pre-processing**

Total dokumen sebelum pre-processing	132516
--------------------------------------	--------



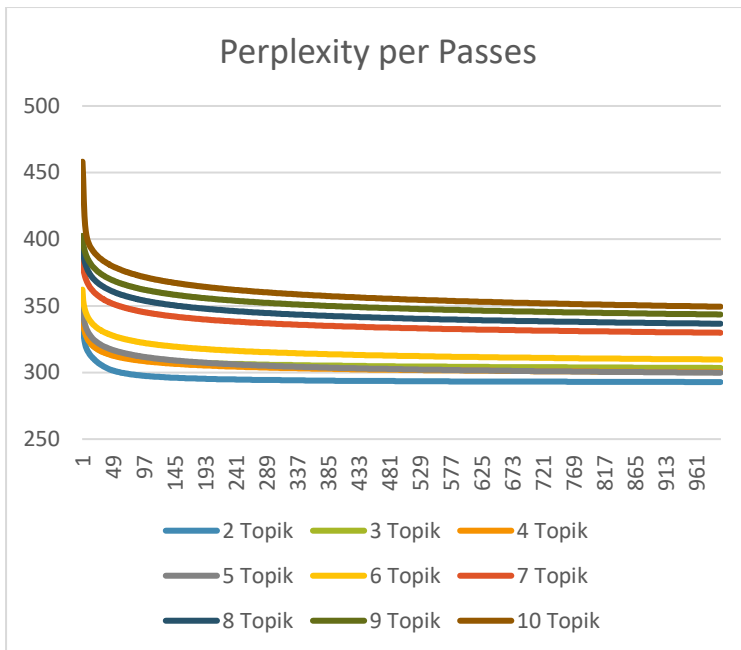
Total dokumen setelah pre-processing	80656
Total token unik	3239
Total author	14

### **6.3 Permodelan Topik dan Validasi**

Dalam pembuatan model topik, yang menjadi fokus utama adalah pencatatan dan pengolahan hasil perplexity yang didapatkan setiap melakukan pembuatan model.

#### **6.3.1 Permodelan Topik dengan Author-Topic Model**

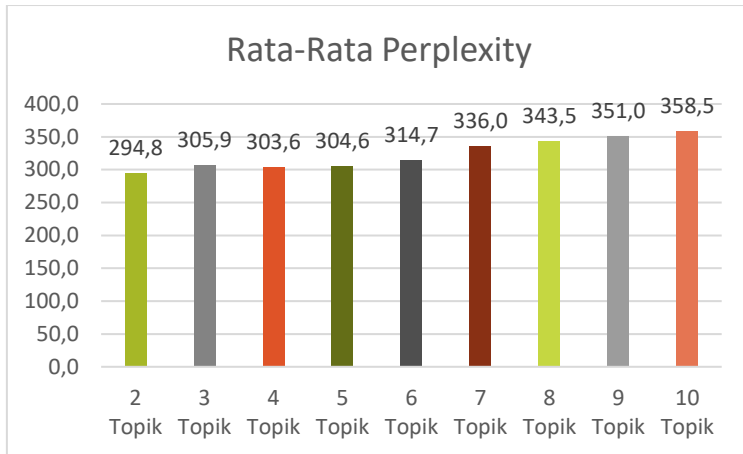
Pada penelitian untuk menentukan model topik berdasarkan author partai, dilakukan 10 kali percobaan pada jumlah topik dari 1 hingga 10 untuk mencari berapa jumlah topik yang baik. Percobaan tersebut dilakukan dengan parameter passes dan jumlah iterasi yang sama, hanya dilakukan perubahan pada jumlah topik.



**Grafik 6.1 Perbandingan Perplexity pada tiap Passes (ATM)**

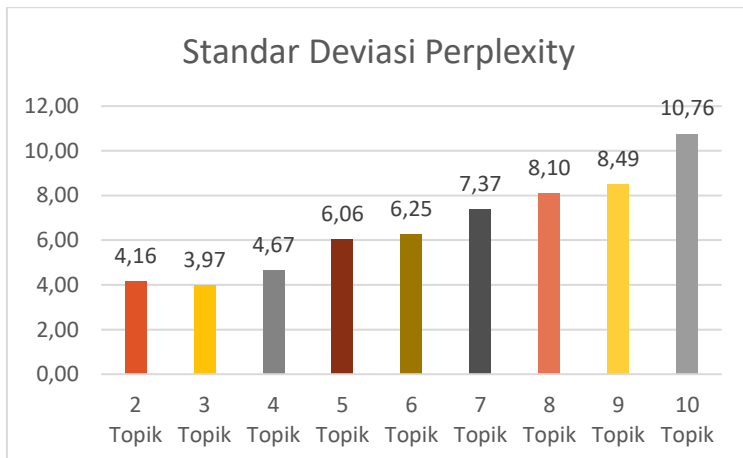
Jumlah passes yang digunakan adalah 1000, karena pada gambar berikut ditunjukkan bahwa model mulai menunjukkan stabilitas pada passes ke 1000. Selain itu, jumlah passes yang semakin besar membuat model tersebut menjadi lebih stabil karena pembentukannya diulang sejumlah dengan parameter passes yang sudah dimasukkan.

Setelah dipastikan jumlah passes yang digunakan adalah 1000, selanjutnya dilakukan percobaan membuat model topik dengan jumlah 2 sampai 10 topik. Selama pembentukan model tersebut, dilakukan juga pencatatan perplexity pada tiap jumlah topik yang dibuat. Dari pencatatan perplexity, selanjutnya dilakukan perhitungan rata-rata, standar deviasi, nilai min dan nilai maksimal perplexity pada setiap model topik yang dibuat.



**Grafik 6.2 Rata-Rata Perplexity ATM**

Grafik 6.2 menunjukkan rata-rata perplexity pada jumlah topik 2 hingga 10. Semakin kecil jumlah topik, semakin kecil juga tingkat keberagaman yang ditunjukkan pada model topik tersebut. Hal ini juga ditunjukkan oleh nilai standar deviasi dari *perplexity* yang dibentuk pada setiap jumlah topik, yang ditunjukkan pada Grafik 6.3.



**Grafik 6.3 Standar Deviasi Perplexity ATM**

## Analisis Penentuan Jumlah Topik

Penentuan jumlah topik dilakukan dengan membandingkan keseluruhan hasil rata-rata, standar deviasi, nilai minimal, dan nilai maksimal dari *perplexity* pada setiap model topik yang sudah dibuat. Jumlah topik yang dipilih adalah topik dengan nilai rata-rata, standar deviasi, nilai minimal, dan nilai maksimal yang berada pada rentang pertengahan dari semua model topik yang dibuat. Perbandingan nilai tersebut ditunjukkan pada Tabel 6.4.

**Tabel 6.4 Perbandingan Perplexity pada tiap Model Topik (ATM)**

	2 Topik	3 Topik	4 Topik	5 Topik	6 Topik	7 Topik	8 Topik	9 Topik	10 Topik
Rata-rata	294,8	305,9	303,6	304,6	314,7	336,0	343,5	351,0	358,5
Stdev	4,16	3,97	4,67	6,06	6,25	7,37	8,10	8,49	10,76
Max	341	344,7	344,7	354,7	362,2	385,6	396,1	402,6	458,2
Min	292,8	303,4	300,3	299,8	309,7	329,8	336,6	343,5	349,4

Dari Tabel 6.4, diketahui bahwa model topik yang memiliki persebaran nilai yang berada pada rentang minimal adalah model dengan jumlah topik 3 dan jumlah topik 4. Hal ini diketahui dengan melihat nilai standar deviasi dari model dengan 3 topik adalah yang terendah dari keseluruhan model, dan juga rata-rata dari model dengan 4 topik yang lebih rendah dari model dengan 3 topik. Model dengan 5 hingga 10 topik tidak dipertimbangkan lebih lanjut karena jarak nilai standar deviasi yang terlalu besar dengan model 4 topik.

Untuk menentukan model topik mana yang akan dipilih, dilakukan juga analisis pada setiap topik yang terbentuk dengan melihat kata-kata yang menyusun topik tersebut. Model dengan 3 topik memiliki detail kata-kata dan probabilitas *author* sebagai berikut.

**Tabel 6.5 Hasil Topik 0 dan Topik 1 pada Model dengan 3 Topik (ATM)**

Topik 0		Topik 1	
Kata	Prob.	Kata	Prob.
partai	0,002095	hukum	0,001205

percaya	0,001197	tantang	0,001205
bangsa	0,001197	pesta	0,001205
peduli	0,001197	ma	0,001205
daerah	0,001197	bijak	0,001205
muda	0,001197	aksi	0,001205
dpd	0,001197	musuh	0,001205
profesional	0,001197	tetangga	0,001205
sapa	0,001197	studi	0,001205
kepala	0,001197	beras	0,001205
<b>Partai</b>	<b>Prob.</b>	<b>Partai</b>	<b>Prob.</b>
NASDEM	0,011102	PDIP	0,991713
PSI	0,992336	PPP	0,991713
PERINDO	0,996142	PAN	0,986286
HANURA	0,993852		
GERINDRA	0,992336		
PKS	0,995595		
BERKARYA	0,994868		

**Tabel 6.6 Hasil Topik 2 pada Model dengan 3 Topik (ATM)**

<b>Topik 2</b>	
<b>Kata</b>	<b>Prob.</b>
bangsa	0,003051
aksi	0,001220
menang	0,001220
cedera	0,001220

hajar	0,001220
demo	0,001220
dunia	0,001220
polisi	0,001220
kutuk	0,001220
serikat	0,001220
<b>Partai</b>	<b>Prob.</b>
PKB	0,994253
GOLKAR	0,991418
NASDEM	0,983061
PKPI	0,993116
DEMOKRAT	0,996540

Kemudian untuk hasil kata-kata dan juga probabilitas partai pada model dengan 4 topik ditunjukkan pada Tabel 6.7 dan Tabel 6.8.

**Tabel 6.7 Hasil Topik 0 dan Topik 1 pada Model dengan 4 Topik (ATM)**

Topik 0		Topik 1	
Kata	Prob.	Kata	Prob.
percaya	0,002686	bangsa	0,005091
aksi	0,002686	upaya	0,001497
partai	0,001492	libat	0,001497
hukum	0,001492	negara	0,001497
nasional	0,001492	muhammadiyah	0,001497
sapa	0,001492	ajak	0,001497
amanat	0,001492	serikat	0,001497

gerak	0,001492	dunia	0,001497
indonesia	0,001492	muda	0,001497
saudara	0,001492	dpd	0,001497
<b>Partai</b>	<b>Prob.</b>	<b>Partai</b>	<b>Prob.</b>
PDIP	0,990631	NASDEM	0,977055
GOLKAR	0,988331	PERINDO	0,994109
NASDEM	0,012145	PKPI	0,990625
HANURA	0,990631	DEMOKRAT	0,995279
GERINDRA	0,988331	PAN	0,984522
PKS	0,993281		

**Tabel 6.8 Hasil Topik 2 dan Topik 3 pada Model dengan 4 Topik (ATM)**

Topik 2		Topik 3	
Kata	Prob.	Kata	Prob.
rapat	0,001525	polisi	0,001523
episode	0,001525	demo	0,001523
baris	0,001525	sadis	0,001523
raih	0,001525	kutuk	0,001523
impor	0,001525	hajar	0,001523
bijak	0,001525	kriminalisasi	0,001523
beras	0,001525	berani	0,001523
presiden	0,001525	kepala	0,001523
gaung	0,001525	oknum	0,001523
perintah	0,001525	tegak	0,001523
<b>Partai</b>	<b>Prob.</b>	<b>Partai</b>	<b>Prob.</b>

PKB	0,989871	PPP	0,987877
PSI	0,984896	BERKARYA	0,989873

Dari hasil topik pada model dengan 4 topik, dapat dilihat bahwa model tersebut memiliki hasil topik yang lebih luas jika dibandingkan dengan model 3 topik. Hal ini dapat dilihat dari kata-kata penyusun tiap topik pada model dengan 4 topik. Dengan kemampuan topik 4 yang lebih baik dalam memetakan hasil topik, maka model dengan jumlah 4 topik adalah model yang dipilih untuk merepresentasikan topik apa saja yang dibicarakan oleh caleg dari 14 partai peserta Pemilu 2019 di Twitter.

### **Dasar Penamaan Kategori Topik dengan ATM**

Dari 4 topik yang dihasilkan dapat diketahui kata-kata yang menyusun setiap topik memiliki keterkaitan satu sama lain. Keempat topik tersebut kemudian diberi label kategori untuk penamaan pada masing-masing topik.

Penamaan topik dilakukan dengan:

1. Melihat 5 kata yang memiliki probabilitas tertinggi pada suatu topik. Hal ini dilakukan karena semakin besar nilai probabilitas menunjukkan tingkat konsistensi kata tersebut dalam membentuk suatu topik.
2. Kata-kata tersebut kemudian digabungkan untuk mencari kemungkinan keterkaitan antar kata untuk muncul pada *tweet* sesuai dengan konteks *tweet* yang dibuat oleh para caleg.
3. Penamaan kategori topik juga didasarkan sesuai dengan kondisi dan waktu pada saat penelitian dilakukan, yang pada tugas akhir ini dilaksanakan pada saat masa Pemilu 2019, sehingga penamaan kategori topik juga disesuaikan dengan kondisi selama Pemilu 2019.

Pada topik 0, terdapat kata ‘percaya’, ‘aksi’, ‘partai’, ‘hukum’, dan ‘nasional’ yang merupakan 5 kata dengan probabilitas tertinggi. Dari kata-kata tersebut, ditemukan keterkaitan dari



kata yang dibahas adalah mengenai hal-hal yang berhubungan dengan hukum, dengan kata pendukung yang penting adalah ‘aksi’ dan ‘hukum’.

Kemudian pada topik 1, 5 kata yang memiliki probabilitas tertinggi antara lain ‘bangsa’, ‘upaya’, ‘libat’, ‘negara’, dan ‘muhammadiyah’. Dari kelima kata tersebut terdapat keterkaitan yang mungkin untuk dibahas pada topik tersebut adalah membahas mengenai isu mengenai pergerakan nasional.

Pada topik 2, juga diambil 5 kata dengan probabilitas tertinggi, antara lain ‘rapat’, ‘episode’, ‘baris’, ‘raih’, dan ‘impor’. Dari kata-kata tersebut kemudian dapat ditemukan keterkaitan kemungkinan hal yang dibahas pada topik tersebut adalah mengenai kebijakan pemerintah, salah satu contohnya yang berkaitan dengan impor.

Lalu pada topik 3 membahas isu kriminalitas, yang didapatkan dari keterkaitan pada 5 kata dengan probabilitas tertinggi antara lain ‘polisi’, ‘demo’, ‘sadis’, ‘kutuk’, dan ‘hajar’. Besar probabilitas tiap kata pada topik 3 memiliki nilai yang sama, sehingga semua kata memiliki peran yang sama dalam membentuk topik tersebut.

Dari hasil tersebut, kemudian dibuat kategori topik sesuai dengan hasil model topik yang dipilih yaitu model dengan 4 topik.

**Tabel 6.9 Kategori Topik (ATM)**

<b>Id_topik</b>	<b>Kategori</b>
0	Isu Hukum
1	Pergerakan Nasional
2	Kebijakan Pemerintah
3	Isu Keamanan

Kategori tersebut selanjutnya digunakan untuk memudahkan visualisasi hasil topik pada Tableau. Hasil topik per partai kemudian ditampilkan menggunakan bentuk lingkaran dengan

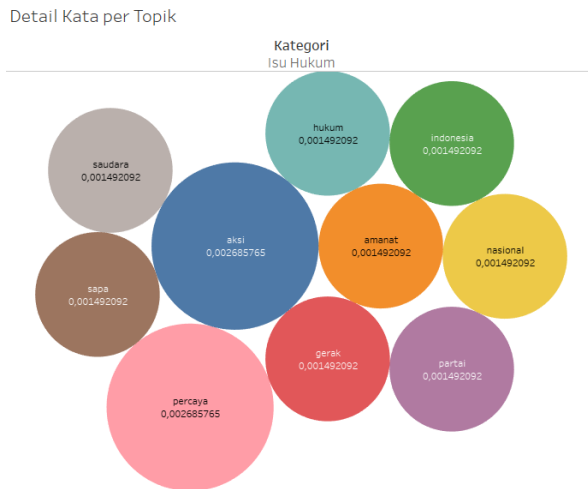
besar diameter sesuai dengan probabilitas kata pada topik tersebut. Untuk mempermudah dalam melihat visualisasi setiap hasil kategori topik yang dibuat menggunakan ATM, dibuat opsi untuk memilih kategori topik yang diinginkan.

#### Kategori

- (All)
- Isu Hukum
- Isu Keamanan
- Kebijakan Pemerintah
- Pergerakan Nasional

**Gambar 6.4 Radio Button untuk memilih Kategori Topik**

Seperti pada contoh berikut, jika opsi kategori yang dipilih adalah ‘Isu Hukum’, maka akan menampilkan detail kata-kata yang menyusun topik tersebut.



**Gambar 6.5 Detail Kata pada Topik Isu Hukum**

Pada kategori topik ‘Isu Hukum’, diketahui terdapat 2 kata yang memiliki probabilitas tertinggi yaitu ‘aksi’ dan ‘percaya’. Dua kata tersebut merupakan kata yang paling signifikan dalam membentuk topik yang pertama, dan 8 kata lainnya memiliki probabilitas yang perbedaannya tidak terlalu signifikan.

Beberapa contoh tweet yang memuat kata-kata yang ada pada topik 'Isu Hukum' antara lain:

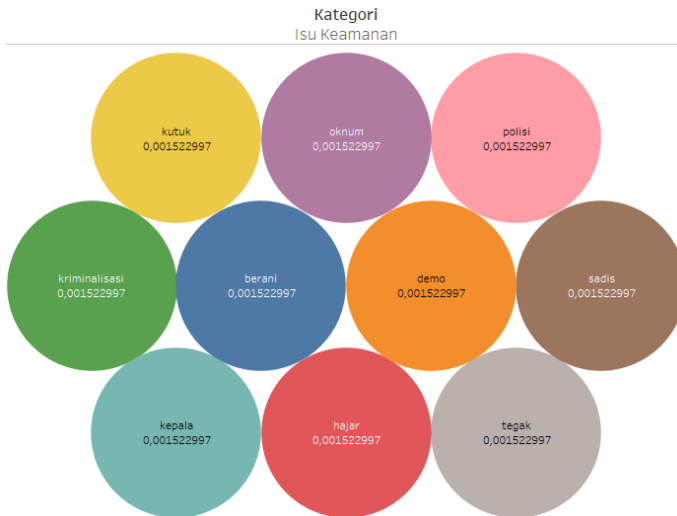
**Tabel 6.10 Contoh Tweet pada Topik Isu Hukum**

Partai	Tweet
PDIP	RT @GesurilD: Masyarakat Tidak <b>Percaya</b> Jokowi Komunis dan Anti Islam  <a href="https://t.co/zduV2H7GVf">https://t.co/zduV2H7GVf</a> #pdiperjuangan"
Golkar	TKN @Jokowi -Ma'ruf Amin mengancam <b>aksi</b> sosialisasi berisi kampanye hitam terhadap pasangan 01. TKN akan menelusuri ibu-ibu yang video sosialisasinya viral di media sosial.  Ace menegaskan isu yang dibawa ibu-ibu dalam video tersebut sepenuhnya hoax.  <a href="https://t.co/msWR2T86IM">https://t.co/msWR2T86IM</a>
Nasdem	Silahkan sj kau laporkan ke ranah <b>hukum</b> Ferdinand Hutahaean @LawanPOLitikJKW. Setidaknya laporan ke polisi soal yg sama juga sdh masuk menguntai mu. Bersiaplah kau masuk bui juga. <a href="https://t.co/KZJij7dzTV">https://t.co/KZJij7dzTV</a>
Hanura	5 Fakta Pembunuhan Dufi yang Pernah Jadi Wartawan Media <b>Nasional</b> <a href="https://t.co/2GXPntH9jJ">https://t.co/2GXPntH9jJ</a> via @okezonenews
Gerindra	RT @toptobs: Selama saya jd kader <b>partai</b> ti 2008 ge saya ngajaga betul batasan-batasan kapan saya hrs bersikap sebagai pribadi saya yg puny...
PKS	RT @cepisyams: @MardaniAliSera Betul Pak, saat ini malah perlu dipublikasikan terutama oleh yang mengemban <b>amanat</b> . Tph panitia hanya menyam...

Pada Tabel 6.10 dapat diketahui contoh *tweet* yang memuat kata-kata yang terdapat pada topik 'Isu Hukum'. Contoh *tweet* yang diambil adalah *tweet* yang mengandung kata 'Aksi' dan

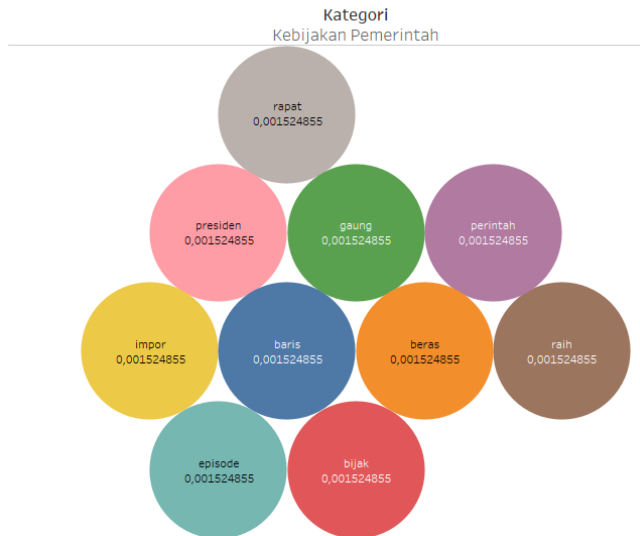
'Hukum'. Tweet pertama dibuat oleh salah satu caleg dari Partai Berkarya, dan *tweet* kedua dibuat oleh caleg dari Partai Nasdem.

Detail Kata per Topik



**Gambar 6.6** Detail Kata pada Topik Isu Keamanan

## Detail Kata per Topik



**Gambar 6.7** Detail Kata pada Topik Kebijakan Pemerintah

Pada kategori topik ‘Isu Keamanan’ dan ‘Kebijakan Pemerintah’, diketahui bahwa semua kata memiliki tingkat probabilitas yang sama, sehingga semua kata pada topik tersebut mempunyai peran yang sama signifikan dalam membentuk topik tersebut. Contoh tweet pada topik ‘Isu Keamanan’ dapat dilihat pada Tabel 6.11.

**Tabel 6.11** Contoh Tweet pada Topik Isu Keamanan

Partai	Tweet
PSI	Fitnah kayak gini, di masjid lagi, kalau ditangkap jgn koar2 <b>kriminalisasi</b> <a href="https://t.co/bI2IMnpPpY">https://t.co/bI2IMnpPpY</a>
PKB	Selamat pagi. <b>Demo</b> buruh di Morowali bkn demo TKA China. Jg bkn demo menolak TKA China. Demo buruh di Morowali terkait Upah Minimum Sektoral Kab (UMSK) setempat, yg saat ini sdg ditangani otoritas terkait di sana. Jgn termakan hoax. Jangan ikut sebarakan hoax. Waspada! adu domba!

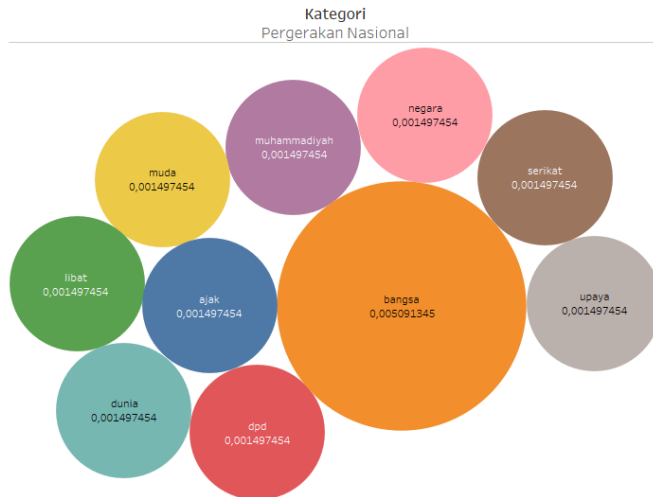
Pada Tabel 6.11, dapat diketahui contoh *tweet* yang dibuat pada topik ‘Isu Keamanan’. *Tweet* pertama mengandung kata ‘kriminalisasi’ dan dibuat oleh caleg dari PSI, dan *tweet* kedua mengandung kata ‘demo’ dan dibuat oleh caleg dari partai PKB.

**Tabel 6.12 Contoh Tweet pada Topik Kebijakan Pemerintah**

Partai	Tweet
PDIP	RT @mohmahfudmd: Kata Kabulog Budi Waseso (Buwas), “Kita tak perlu <b>impor beras</b> , gudang Bulog msh penuh, tak mampu menampung beras impor”.
PAN	UU Pemilu No.7 Tahun 2017 Pasal 170 tidak ada <b>perintah Presiden</b> harus mundur bila maju Capres... Semoga kita tidak terjebak provokasi yang tidak bertanggung jawab...  Kritis itu cerdas... <a href="https://t.co/7rqu9Vwac3">https://t.co/7rqu9Vwac3</a>

Kemudian pada Tabel 6.12 dapat diketahui contoh *tweet* yang membahas topik ‘Kebijakan Pemerintah’. *Tweet* pertama, yang dibuat oleh caleg dari PDIP, mengandung kata ‘impor’ dan ‘beras’. Pada *tweet* kedua mengandung kata ‘perintah’ dan ‘presiden’, dan *tweet* tersebut dibuat oleh caleg dari PAN.

## Detail Kata per Topik



**Gambar 6.8 Detail Kata pada Topik Pergerakan Nasional**

Pada kategori topik ‘Pergerakan Nasional’, kata ‘bangsa’ memiliki tingkat probabilitas yang paling tinggi dibandingkan dengan kata-kata yang lain, kemudian 9 kata yang lain memiliki probabilitas yang sama signifikan satu sama lainnya. Contoh *tweet* yang membahas topik ‘Pergerakan Nasional’ adalah sebagai berikut.

**Tabel 6.13 Contoh Tweet pada Topik Pergerakan Nasional**

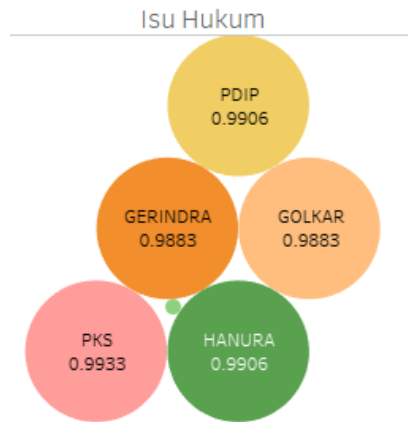
Partai	Tweet
PKS	Pancasila pada diri seorang prajurit pembela <b>bangsa</b> jenderal bintang tiga, tentu tidak diragukan lagi, nasionalisme @Prabowo & @sandiuono juga telah teruji. Mereka 2019 tetap Pancasila & jaga NKRI. Gelora #2019GantiPresiden adalah memilih pemimpin baru yg tdk akan gadaikan negeri
Golkar	Ralat: Dalam <b>upaya</b> pemberantasan korupsi ini, Pak @jokowi mengatakan aspek pencegahan yang sangat penting, bukan hanya aspek penindakan. Pak Jokowi menawarkan agenda aksi yang konkret

	dan komprehensif. Pak Jokowi sudah meletakkan fondasi berupa Strategi Nasional Pencegahan Korupsi
--	---

Tweet pertama dibuat oleh salah satu caleg yang berasal dari PKS, yang dalam *tweetnya* mengandung kata ‘bangsa’. Pada *tweet* kedua terdapat kata ‘upaya’, dan *tweet* tersebut dibuat oleh caleg dari Partai Golkar.

### Analisis Author tiap Topik

Dari permodelan topik yang sudah dibuat, dapat diketahui kemungkinan partai sebagai *author* berpartisipasi dalam membentuk suatu topik. Dengan 4 topik yang sudah terbentuk, didapatkan probabilitas 14 partai dalam membahas suatu topik selama Pemilu 2019.



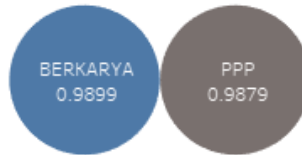
**Gambar 6.9 Detail Author Partai pada Topik Isu Hukum**

Pada kategori topik ‘Isu Hukum’, dapat diketahui bahwa caleg partai yang memiliki probabilitas tertinggi dalam membahas topik tersebut adalah PKS, kemudian disusul oleh Hanura, PDIP, Gerindra, Golkar, dan Nasdem yang ditunjukkan dengan lingkaran terkecil. Dalam topik ini, diketahui terdapat 6 partai yang mempunyai kemungkinan dalam membicarakan topik ‘Isu Hukum’.



### Isu Keamanan

---

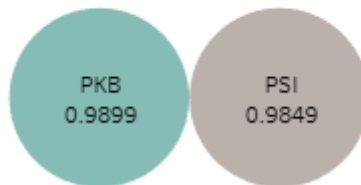


**Gambar 6.10 Detail Author Partai pada Topik Isu Keamanan**

Pada topik ‘Isu Keamanan’ didominasi oleh 2 partai yaitu Berkarya dan PPP. Caleg dari kedua partai tersebut memiliki peluang dalam membicarakan topik mengenai isu keamanan pada *tweet* yang dibagikan.

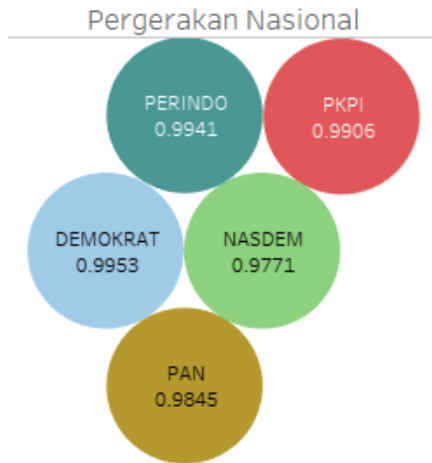
### Kebijakan Pemerintah

---



**Gambar 6.11 Detail Author Partai pada Topik Kebijakan Pemerintah**

Pada kategori topik ‘Kebijakan Pemerintah’, terdapat 2 partai yang memiliki kemungkinan untuk membahas topik tersebut, yaitu Partai Kebangkitan Bangsa (PKB) dan Partai Solidaritas Indonesia (PSI). Kemungkinan PKB dalam membahas topik tersebut sedikit lebih besar jika dibandingkan dengan probabilitas dari PSI.

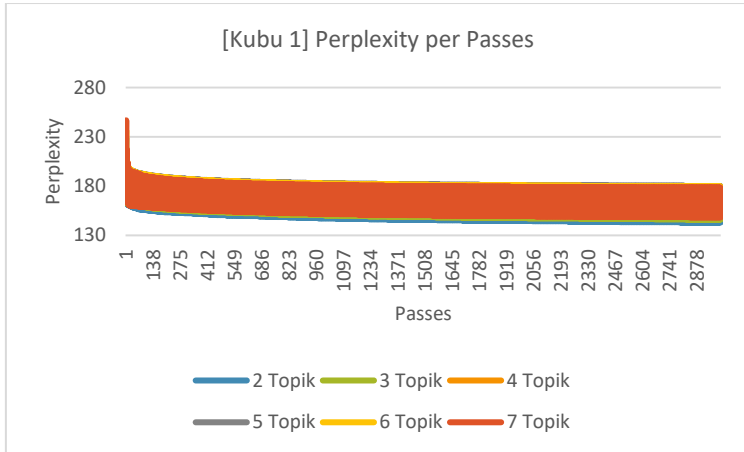


**Gambar 6.12 Detail Author Partai pada Topik Pergerakan Nasional**

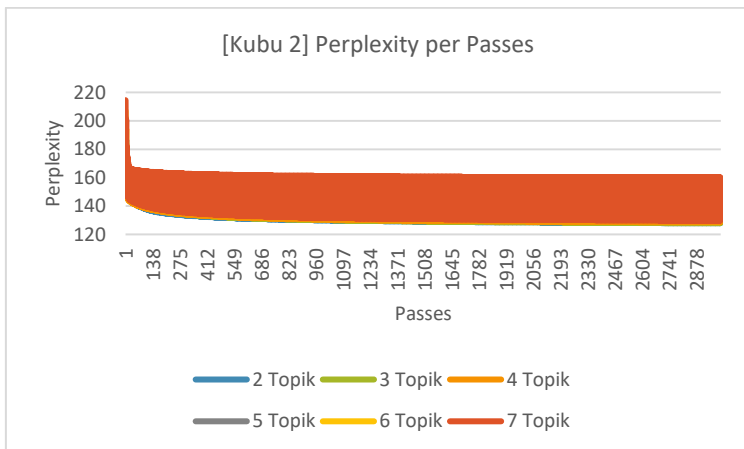
Pada kategori topik ‘Pergerakan Nasional’ memiliki kemungkinan dibahas oleh 4 partai, dengan urutan probabilitas tertinggi adalah Partai Demokrat, kemudian disusul oleh Perindo, PKPI, PAN, dan kemudian Partai Nasdem.

### 6.3.2 Permodelan Topik dengan Latent Dirichlet Allocation

Pada penelitian untuk menentukan model topik berdasarkan author partai, dilakukan 7 kali percobaan pada jumlah topik dari 1 hingga 7 untuk mencari berapa jumlah topik yang baik. Percobaan tersebut dilakukan dengan parameter passes dan jumlah iterasi yang sama, hanya dilakukan perubahan pada jumlah topik.

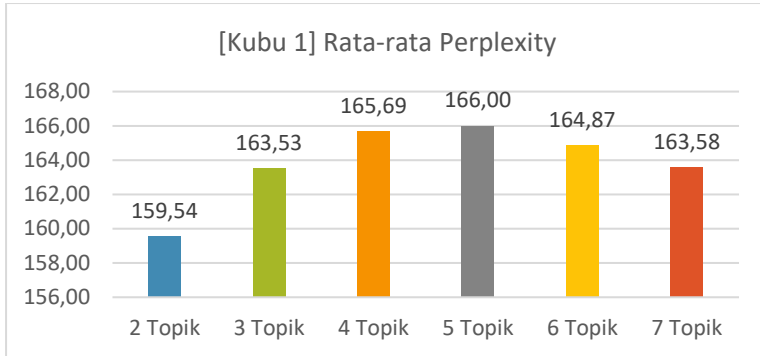


**Grafik 6.4 Perbandingan Perplexity pada tiap Passes (LDA-Kubu 01)**

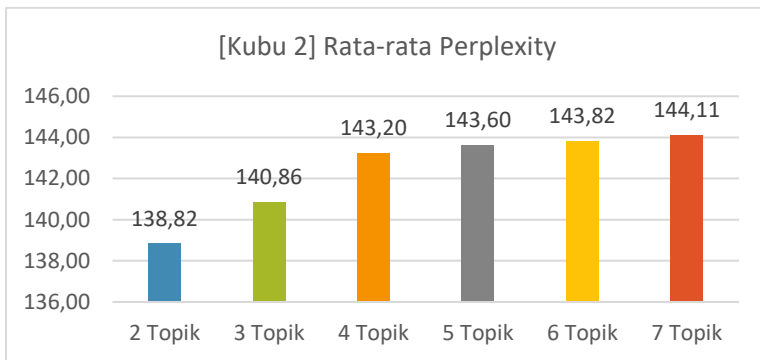


**Grafik 6.5 Perbandingan Perplexity pada tiap Passes (LDA-Kubu 02)**

Pembentukan model dilakukan dengan menggunakan 1000 *passes* dan 10000 iterasi, hal ini dilakukan untuk membentuk model yang lebih stabil, yang ditunjukkan pada Grafik 6.4 dan Grafik 6.5. Dengan melihat tingkat kestabilan *perplexity* tersebut, selanjutnya dilakukan perhitungan untuk mengetahui rata-rata, standar deviasi, nilai minimal, dan nilai maksimal pada masing-masing kubu capres dan cawapres.

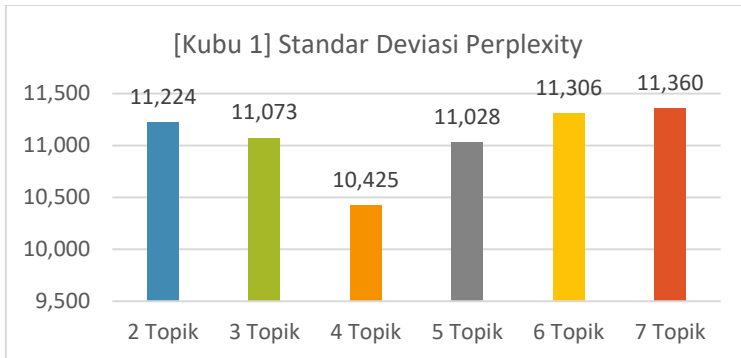


**Grafik 6.6 Rata-Rata Perplexity LDA-Kubu 01**

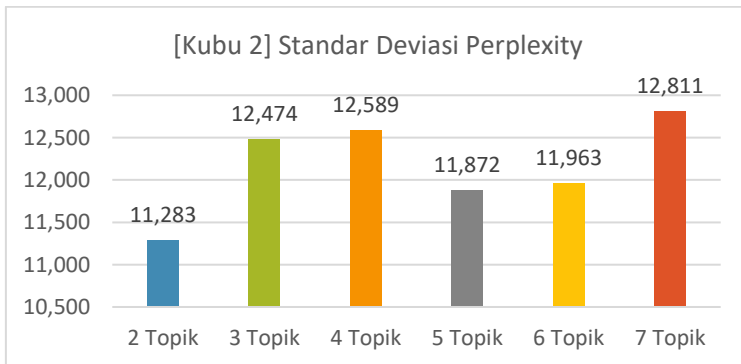


**Grafik 6.7 Rata-Rata Perplexity LDA-Kubu 02**

Grafik 6.6 dan Grafik 6.7 menunjukkan rata-rata perplexity pada jumlah topik 2 hingga 7 pada masing-masing kubu capres dan cawapres. Semakin kecil jumlah topik, semakin kecil juga tingkat keberagaman yang ditunjukkan pada model topik tersebut. Hal ini juga ditunjukkan oleh nilai standar deviasi dari *perplexity* yang dibentuk pada setiap jumlah topik, yang ditunjukkan pada Grafik 6.8 dan Grafik 6.9.



**Grafik 6.8 Standar Deviasi Perplexity LDA-Kubu 01**



**Grafik 6.9 Standar Deviasi Perplexity LDA-Kubu 02**

### Analisis Penentuan Jumlah Topik

Penentuan jumlah topik pada tiap kubu dilakukan setelah perhitungan untuk mengetahui rata-rata, standar deviasi, nilai minimal, dan nilai maksimal *perplexity* pada setiap model topik didapatkan.

**Tabel 6.14 Perbandingan Perplexity pada tiap Model Topik (LDA-Kubu 01)**

KUBU 01	2 Topik	3 Topik	4 Topik	5 Topik	6 Topik	7 Topik
Rata-rata	159,54	163,53	165,69	166,00	164,87	163,58
Stdev	11,22	11,07	10,42	11,03	11,31	11,36

Maks	201,9	209,5	217,8	228,9	236,8	246,9
Min	143	146,1	151,6	150,6	149,8	148,3

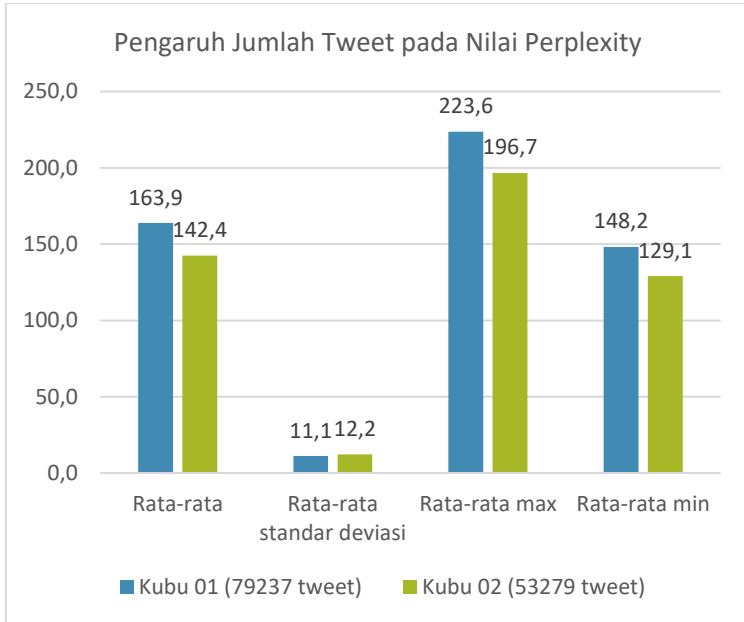
Pada Tabel 6.14, menunjukkan bagaimana persebaran nilai rata-rata, standar deviasi, nilai minimal dan nilai maksimal dari permodelan topik yang dibuat dengan menggabungkan seluruh *tweet* dari caleg kubu Jokowi-Ma'ruf. Dari data tersebut, terdapat 2 model topik yang termasuk pada area terkecil atau ditandai dengan warna hijau-kuning, yaitu model dengan 2 topik dan 3 topik. Sementara untuk model dengan 4 sampai 7 topik memiliki nilai *perplexity* yang relatif tinggi, dan hal tersebut menyebabkan model tersebut tidak dipertimbangkan lebih lanjut.

**Tabel 6.15 Perbandingan Perplexity pada tiap Model Topik (LDA-Kubu 02)**

KUBU 02	2 Topik	3 Topik	4 Topik	5 Topik	6 Topik	7 Topik
Rata-rata	138,82	140,86	143,20	143,60	143,82	144,11
Stdev	11,28	12,47	12,59	11,87	11,96	12,81
Maks	173,4	186,8	195,6	202,6	207,1	214,6
Min	127,9	127,9	128,5	130,8	130	129,7

Pada kubu 02, didapatkan hasil yang hampir sama dengan permodelan topik yang dilakukan pada kubu 01. Tabel 6.15 menunjukkan model dengan 2 dan 3 topik memiliki nilai *perplexity* yang relatif rendah jika dibandingkan dengan model dengan jumlah topik 4 sampai 7.

Dari 2 percobaan yang dilakukan, didapatkan nilai *perplexity* pada tiap topik yang berbeda pada kedua kubu. Hal ini diketahui dipengaruhi oleh jumlah *tweet* dari kubu 01 yang lebih banyak jika dibandingkan dengan *tweet* dari kubu 02. Perbandingan besar nilai *perplexity* dari kedua kubu dapat dilihat pada Grafik 6.10.



**Grafik 6.10 Pengaruh Jumlah Tweet pada Nilai Perplexity**

Dalam menentukan model topik mana yang dipilih untuk merepresentasikan kedua kubu capres dan cawapres, selanjutnya dilakukan analisis untuk melihat rincian kata-kata yang membentuk topik-topik tersebut. Meskipun model topik 2 memiliki nilai *perplexity* yang terendah, hal itu tidak langsung menjadikan model tersebut adalah model yang dipilih. Semakin kecil jumlah pengelompokan maka semakin kecil juga nilai *perplexity*-nya, sehingga perlu juga dilakukan analisis pada model topik lain yang masih berada pada batas dimana model tersebut masih terhitung memiliki nilai *perplexity* yang relatif rendah.

Hasil permodelan topik yang akan dibahas terlebih dahulu adalah permodelan topik yang dilakukan untuk mengetahui topik bahasan dari kubu 01. Berdasarkan analisis hasil *perplexity* yang dilakukan, maka model topik yang dipertimbangkan untuk dipilih adalah model dengan 2 dan 3

topik. Tabel 6.16 menunjukkan model dengan 2 topik dari kubu 01.

**Tabel 6.16 Hasil Topik 0 dan Topik 1 pada Model dengan 2 Topik (LDA-Kubu 01)**

Topik 0		Topik 1	
Kata	Prob.	Kata	Prob.
jokowi	0,018810	indonesia	0,032346
politik	0,010389	kerja	0,014013
prabowo	0,008714	juang	0,009809
dukung	0,008558	jalan	0,009790
suara	0,008029	dpr	0,009655
menang	0,007708	presiden	0,008345
rakyat	0,007427	bangsa	0,008070
agama	0,007390	masyarakat	0,007982
lihat	0,006684	teman	0,006959
nyata	0,006607	desa	0,006681
negara	0,006470	muda	0,006342
partai	0,006163	bantu	0,006297
hoax	0,005928	dunia	0,006143
hasil	0,005231	partai	0,005957
baca	0,005152	keluarga	0,005923

Kemudian Tabel 6.17 dan Tabel 6.18 menunjukkan detail kata-kata topik pada model dengan 3 topik.

**Tabel 6.17 Hasil Topik 0 dan Topik 1 pada Model dengan 3 Topik (LDA-Kubu 01)**

Topik 0	Topik 1
---------	---------



Kata	Prob.	Kata	Prob.
dukung	0,017003	indonesia	0,044312
partai	0,015048	jokowi	0,030690
politik	0,014007	kerja	0,019198
juang	0,011105	presiden	0,017185
suara	0,010829	jalan	0,013415
menang	0,010418	dpr	0,013226
negara	0,010359	hasil	0,011777
rakyat	0,010019	masyarakat	0,010936
agama	0,009963	lihat	0,010910
bangsa	0,009137	perintah	0,009177
nyata	0,008908	desa	0,009152
dunia	0,006955	muda	0,008689
nama	0,006833	buka	0,008623
hidup	0,006713	dapil	0,007550
keluarga	0,006705	kampanye	0,007233

**Tabel 6.18 Hasil Topik 2 pada Model dengan 3 Topik (LDA-Kubu 01)**

Topik 2	
Kata	Prob.
prabowo	0,017964
hoax	0,012220
teman	0,012050
bantu	0,010900
baca	0,010621

bohong	0,008276
lawan	0,007956
debat	0,007579
korupsi	0,007405
percaya	0,007176
kalah	0,007135
fitnah	0,006876
buah	0,006867
sebar	0,006845
data	0,006634

Dari detail topik tersebut, dapat dilihat perbedaan detail kata pada model dengan 2 topik dan model dengan 3 topik. Dengan pertimbangan nilai *perplexity* dan kemampuan model tersebut dalam mengumpulkan kata-kata dalam suatu topik, maka model topik yang akan merepresentasikan bahasan pada caleg dari kubu 01 adalah model dengan 3 topik.

Pada kubu 02, juga akan dilakukan perbandingan dari 2 model dengan nilai *perplexity* yang relatif rendah yaitu model dengan 2 topik dan model dengan 3 topik. Tabel 6.19 menunjukkan detail pada model topik dengan 2 topik.

**Tabel 6.19 Hasil Topik 0 dan Topik 1 pada Model dengan 2 Topik (LDA-Kubu 02)**

Topik 0		Topik 1	
Kata	Prob.	Kata	Prob.
prabowo	0,024462	rakyat	0,012428
indonesia	0,018604	jokowi	0,010933
menang	0,016573	negara	0,010097
partai	0,015313	kerja	0,009632

dukung	0,015299	politik	0,009209
suara	0,013786	kpu	0,009040
masyarakat	0,011066	perintah	0,008740
sandi	0,010125	presiden	0,008089
dpr	0,009504	hasil	0,007399
juang	0,009071	nyata	0,006656
bangsa	0,007140	curang	0,006573
kampanye	0,007123	data	0,006486
jalan	0,006855	lihat	0,006442
rakyat	0,006563	indonesia	0,005751
teman	0,006434	hukum	0,005517

Pada Tabel 6.20 dan Tabel 6.21 memuat detail model dengan 3 topik dari kubu 02.

**Tabel 6.20 Hasil Topik 0 dan Topik 1 pada Model dengan 3 Topik (LDA-Kubu 02)**

Topik 0		Topik 1	
Kata	Prob.	Kata	Prob.
partai	0,025096	prabowo	0,028033
masyarakat	0,018147	menang	0,018990
politik	0,017140	dukung	0,017533
dpr	0,015576	suara	0,015802
teman	0,010545	jokowi	0,014224
bogor	0,010395	presiden	0,013089
bantu	0,010239	kpu	0,011761
kader	0,009858	sandi	0,011602

jawa	0,009810	hasil	0,009633
sehat	0,009791	nyata	0,008660
saudara	0,009718	curang	0,008551
rawan	0,009501	data	0,008438
keluarga	0,009140	lihat	0,008385
salam	0,008807	kampanye	0,008164
baca	0,008729	hukum	0,007177

**Tabel 6.21 Hasil Topik 2 pada Model dengan 3 Topik (LDA-Kubu 02)**

Topik 2	
Kata	Prob.
indonesia	0,038503
rakyat	0,031668
negara	0,017566
kerja	0,016760
jalan	0,015651
bangsa	0,015220
perintah	0,015205
juang	0,013899
pimpin	0,013563
adil	0,009684
ekonomi	0,008705
dunia	0,008572
hati	0,008402
tinggal	0,008248

hidup	0,008247
-------	----------

Dari detail topik yang sudah didapatkan untuk model topik pada kubu 02, dipilih model dengan 3 topik untuk merepresentasikan topik apa saja yang dibahas oleh caleg dari kubu 02.

### Dasar Penamaan Kategori Topik dengan LDA

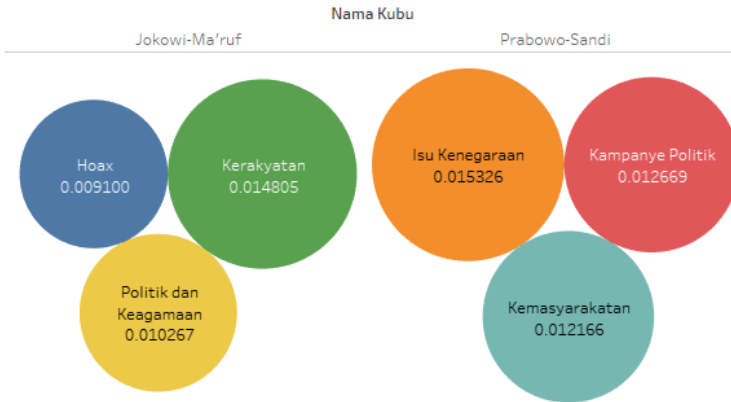
Dari permodelan topik yang dilakukan pada masing-masing kubu capres dan cawapres, perlu dilakukan pemberian kategori pada setiap topik yang sudah dibuat. Pertimbangan dalam pemberian nama kategori topik dilakukan dengan cara yang sama pada saat menentukan nama topik dari permodelan menggunakan ATM pada subbab 6.3.1. Untuk memberikan nama pada topik, dilakukan dengan mengacu pada 5 kata yang memiliki probabilitas tertinggi, kemudian dianalisis kemungkinan keterkaitan dalam interpretasi kelima kata tersebut. Pemberian label kategori ini juga untuk memudahkan visualisasi topik yang dibicarakan oleh caleg dari masing-masing kubu capres dan cawapres. Penamaan kategori tersebut tentunya juga dipengaruhi oleh kata-kata yang menyusun topik tersebut. Kategori topik yang berhasil didapatkan dari masing-masing kubu dapat dilihat pada Tabel 6.22.

**Tabel 6.22 Kategori Topik (LDA)**

Kubu	Id Topik	Kategori topik
01	0	Politik dan Keagamaan
	1	Kerakyatan
	2	Hoax
02	3	Kemasyarakatan
	4	Kampanye Politik
	5	Isu Kenegaraan

Topik yang dibahas oleh masing-masing kubu kemudian dapat divisualisasikan sebagai berikut.

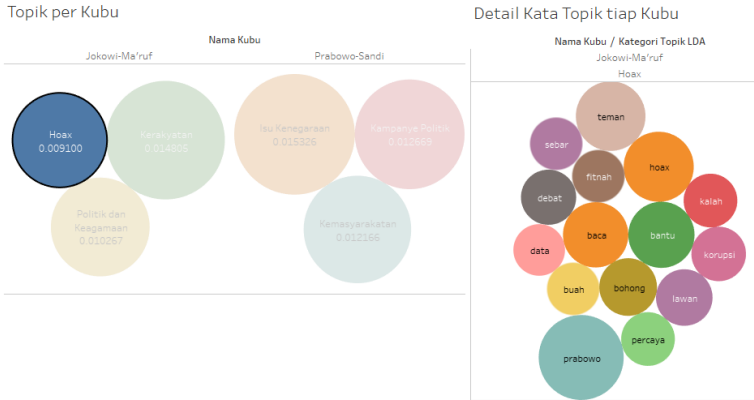
## Topik per Kubu



**Gambar 6.13 Hasil Topik berdasarkan Kubu Capres dan Cawapres (LDA)**

Pada kubu 01, dari 3 topik yang terbentuk dapat diketahui bahwa topik ‘Kerakyatan’ memiliki probabilitas yang lebih tinggi jika dibandingkan dengan topik ‘Politik dan Keagamaan’ dan juga topik ‘Hoax’. Sementara pada kubu 02 diketahui ‘Isu kenegaraan’ menjadi topik yang memiliki kemungkinan tertinggi untuk dibahas oleh caleg kubu tersebut, disusul dengan topik ‘Kampanye Politik’ dan juga topik ‘Kemasyarakatan’.

Dari setiap kategori tersebut kemudian digunakan sebagai opsi yang dapat dipilih dan juga memudahkan proses visualisasi. Visualisasi topik beserta detail kata-kata yang menyusunnya dilakukan dengan memilih topik dan kemudian akan ditampilkan detail dari topik yang dipilih.



**Gambar 6.14 Detail Kata pada Topik Hoax**

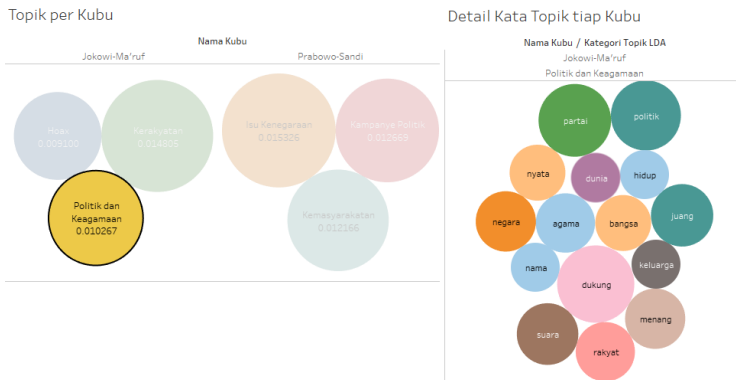
Sebagai contoh, pada Gambar 6.14 ditunjukkan detail topik pada topik 'Hoax' yang merupakan topik dari kubu Jokowi-Ma'ruf. Pada saat topik 'Hoax' dipilih, pada bagian sebelah kanan akan secara otomatis menyesuaikan dan menampilkan detail kata yang menyusun topik 'Hoax'. Topik 'Hoax' didukung oleh beberapa kata seperti 'hoax', 'bohong', 'lawan', dan 'data'. Pada gambar tersebut dapat diketahui detail kata yang mendukung topik tersebut, beserta urutan tingkat probabilitas kata tersebut yang ditunjukkan dengan perbedaan ukuran lingkaran dari masing-masing kata.

**Tabel 6.23 Contoh Tweet pada Topik Hoax**

Partai	Tweet
PPP	<p>Rekam jejak tak lekang oleh waktu...</p> <p>Semakin banyak punggawa2-mu semburkan dusta, <b>fitnah</b> dan kebencian..., semakin banyak pula terkuak rekam jejak kelam-mu...</p> <p>Bisa jadi ada bom waktu tentang hutang2-mu...</p> <p>#01JKWUNTUKMAJU #JokowiMemberikanBukti  <a href="https://t.co/0hEAeEJfnN">https://t.co/0hEAeEJfnN</a></p>

Gerindra	Heran. Paling getol teriak <b>hoax</b> tapi paling getol juga buat hoax. Sampai pasien kanker aja jadi bahan buat bikin hoax dan fitnah ke Sandi. Tega banget cebong
----------	--

Dari Tabel 6.23 dapat diketahui contoh *tweet* yang membahas topik ‘Hoax’. *Tweet* pertama disumbang oleh caleg dari PPP, yang pada *tweet*nya memiliki kata ‘fitnah’. Pada *tweet* kedua dibuat oleh caleg dari Partai Gerindra, dan pada *tweet*nya terdapat kata ‘hoax’.



**Gambar 6.15** Detail Kata pada Topik Politik dan Keagamaan

Kemudian pada topik ‘Politik dan Keagamaan’ diketahui terdapat kata-kata yang mendukung topik tersebut yaitu ‘dukung’, ‘suara’, ‘politik’, ‘partai’, dan ‘agama’. Kata ‘dukung’ memiliki probabilitas yang lebih tinggi jika dibandingkan dengan kata-kata penyusun yang lain. Contoh *tweet* yang membahas Topik Politik dan Keagamaan adalah sebagai berikut.

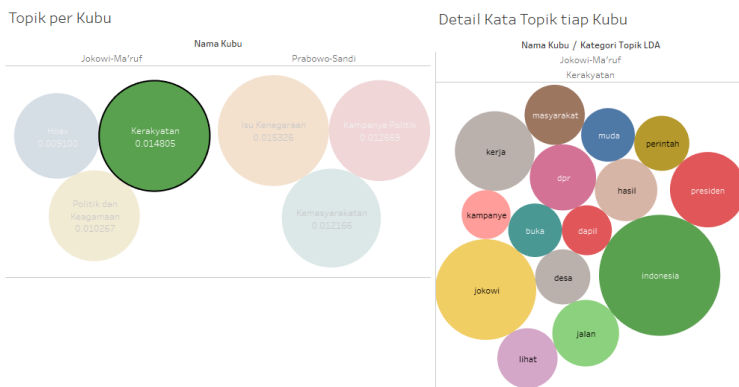
**Tabel 6.24** Contoh *Tweet* pada Topik Politik dan Keagamaan

Partai	Tweet
Perindo	Elektabilitas sering dibicarakan menjelang pemilihan umum. Elektabilitas <b>partai politik</b> berarti tingkat keterpilihan partai politik di publik.



	Elektabilitas partai tinggi berarti partai tersebut memiliki daya pilih... <a href="https://t.co/cufEZqCEKS">https://t.co/cufEZqCEKS</a>
Hanura	@Ferdinand_Haeen @prabowo @sandiuono Ini yg sdh mulai miring antenanya...janji <b>agama</b> pun di lontarkan..ustad ijmak ulama saja tidak berani bilanginya begitu....pesan yg besar lg kaca cermin..biar kamu tau siapa dirimu @Ferdinand_Haean itu di mata ummat islam.

Tabel 6.24 menunjukkan contoh *tweet* yang membahas topik ‘Politik dan Keagamaan’. *Tweet* pertama dibuat oleh caleg dari Perindo, dan pada *tweet*nya terdapat kata ‘politik’. Sementara pada *tweet* kedua dibuat oleh caleg dari Partai Hanura, dan pada *tweet*nya terdapat kata ‘agama’.



**Gambar 6.16 Detail Kata pada Topik Kerakyatan**

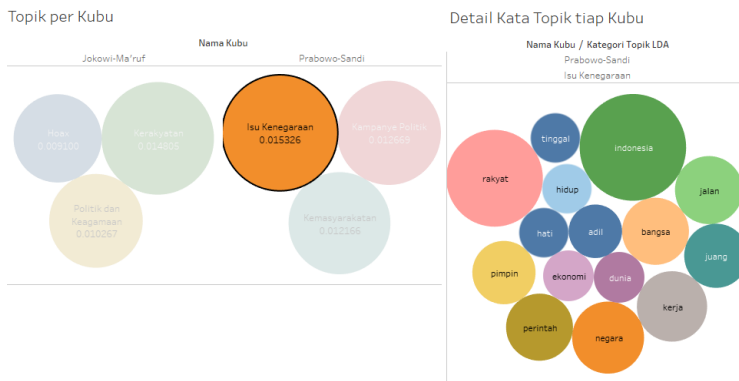
Pada topik ‘Kerakyatan’ terlihat bahwa kata ‘Indonesia’ dan ‘jokowi’ memiliki probabilitas yang lebih tinggi jika dibandingkan dengan kata lainnya. Selain 2 kata tersebut terdapat juga kata-kata pendukung lain yang membentuk topik, seperti ‘masyarakat’, ‘kerja’, ‘desa’, dan ‘muda’.

**Tabel 6.25 Contoh Tweet pada Topik Kerakyatan**

Partai	Tweet
--------	-------

Demokrat	<p>14 Prioritas Demokrat untuk rakyat merupakan komitmen perjuangan apabila Partai Demokrat mendapatkan amanah dari Rakyat pada 17 April 2019.</p> <p>Prioritas 1 : Ciptakan lebih banyak lapangan <b>kerja</b>, termasuk untuk milenial dan perempuan, serta membatasi tenaga kerja asing  <a href="https://t.co/zQBE0kclke">https://t.co/zQBE0kclke</a></p>
Hanura	<p>Bagi <b>masyarakat</b> awam, solusi banjir adalah menunggu air menyusut atau kering #banjirjakarta</p>

Tweet pertama dibuat oleh caleg dari Partai Demokrat, dan pada *tweet* yang dibuat caleg tersebut mengandung kata ‘kerja’. Tweet kedua dibuat oleh caleg dari Partai Hanura, dan pada *tweet* tersebut terdapat kata ‘masyarakat’.



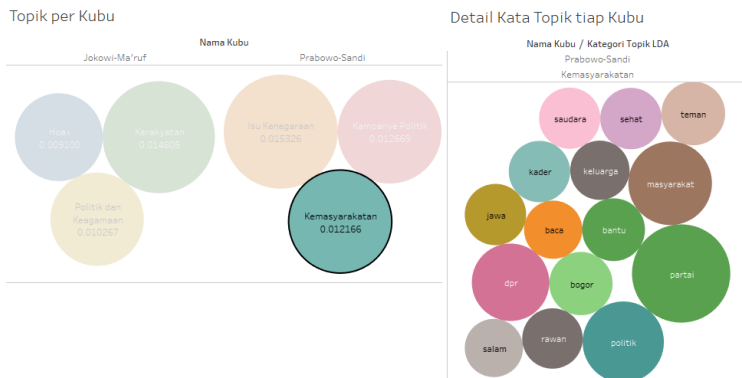
**Gambar 6.17** Detail Kata pada Topik Isu Kenegaraan

Pada kubu 02, terdapat topik ‘Isu Kenegaraan’ yang berisi beberapa kata penting seperti ‘indonesia’, ‘rakyat’, ‘perintah’, ‘negara’, dan ‘kerja’. Dari kata-kata tersebut, terlihat bahwa kata ‘indonesia’ dan ‘rakyat’ memiliki probabilitas yang tinggi jika dibandingkan dengan kata lainnya. Contoh *tweet* yang membahas topik ‘Isu Kenegaraan’ adalah sebagai berikut.

Tabel 6.26 Contoh Tweet pada Topik Isu Kenegaraan

Partai	Tweet
PDIP	Bangga menjadi kader NU. Hari ini Muslimat NU menyelenggarakan Harlah ke-73 di GBK. Muslimat NU ini kokoh menjaga Pancasila dan NKRI melalui pemberdayaan <b>ekonomi</b> , pendidikan, dan pencerahan keagamaan #HarlahMuslimatNU73
PAN	Assalammu'alaikum Wr. Wb.,  Terima kasih kepada teman-teman relawan di Pondok Gede, Pondok Melati, Bekasi Utara yang sudah hadir di acara pelatihan bersama. Saya percaya dari silaturahmi ini, akan menimbulkan semangat <b>juang</b> untuk Kota Bekasi yang maPAN.(1) <a href="https://t.co/IS38uA2i6s">https://t.co/IS38uA2i6s</a>

Tweet pertama yang membahas topik 'Isu Kenegaraan' adalah *tweet* dari salah satu caleg PDIP, dan pada *tweet* tersebut mengandung kata 'ekonomi'. Pada *tweet* kedua dibuat oleh caleg dari PAN, dan pada *tweetnya* mengandung kata 'juang'.



Gambar 6.18 Detail Kata pada Topik Kemasyarakatan

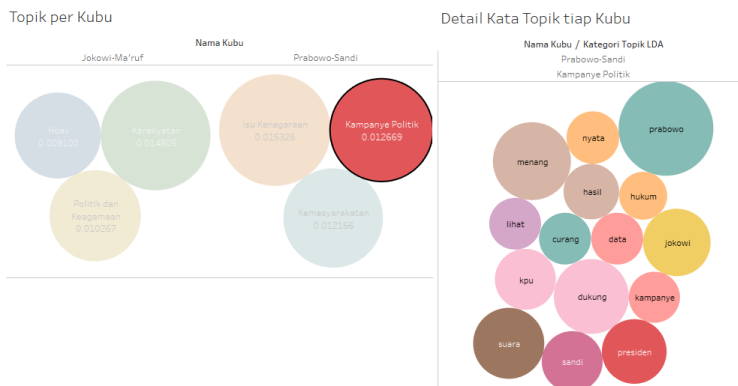
Kemudian pada topik 'Kemasyarakatan' disusun oleh beberapa kata seperti 'partai', 'masyarakat', 'dpr', dan 'kader'. Dari detail topik tersebut, terlihat kata-kata penyusunnya memiliki

probabilitas yang hampir sama besarnya, kecuali kata ‘partai’ yang memiliki probabilitas lebih tinggi. Contoh *tweet* yang membahas topik ‘Isu Kenegaraan’ dapat dilihat pada tabel berikut.

**Tabel 6.27 Contoh Tweet pada Topik Kemasyarakatan**

Partai	Tweet
Golkar	Selamat kepada seluruh lapisan <b>masyarakat</b> Jakarta khususnya dan Indonesia umumnya yang telah resmi memiliki Moda Raya Terpadu atau Masa Rapid Transit MRT #mrt #mrtjakarta #peresmianmrtjakarta #golkarjatim #golkar #golkarsolid #yuliandriyani #indonesia #jawatimur #seputarjawatimur <a href="https://t.co/SyXXgxrTxL">https://t.co/SyXXgxrTxL</a>
Gerindra	Memberikan sosialisasi terkait huruf Braille di kertas suara untuk para penyandang disabilitas (tunanetra). Alat <b>bantu</b> Braille bagi pemilih tunanetra dpt memudahkan pemilih tunanetra dlm mencoblos surat suara. <a href="https://t.co/HnSzhY4yaB">https://t.co/HnSzhY4yaB</a>

Tweet pertama dibuat oleh caleg dari Partai Golkar dan pada *tweet* tersebut terdapat kata ‘masyarakat’. Sementara pada *tweet* kedua dibuat oleh caleg dari Partai Gerindra, yang menyebutkan kata ‘bantu’ pada *tweet*.



**Gambar 6.19 Detail Kata pada Topik Kampanye Politik**

Pada topik ‘Kampanye Politik’ yang dibahas oleh caleg dari kubu 02, terdapat beberapa kata yang menyusun topik tersebut. Kata ‘prabowo’ dan ‘jokowi’ terdapat pada 1 topik yang sama, diperkuat dengan kata ‘suara’, ‘kpu’, ‘kampanye’, ‘dukung’, dan ‘presiden’. Contoh *tweet* yang membahas topik ‘Kampanye Politik’ dapat dilihat pada Tabel 6.28.

**Tabel 6.28 Contoh Tweet pada Topik Kampanye Politik**

Partai	Tweet
PKPI	Ayo kita Meriahkan <b>Kampanye</b> Akbar Jokowi - Amin Tanggal 5 April 2019 di Cirebon dan Indramayu.  #PKPI #Pemilu2019 #dapilJabar8 #cirebon #indramaju #Jokowi2Periode #partaizamanwow #pojokkananbawah20 #01JokowiLagi #koalisiindonesiakerja #TKN #calegdprii #pemilu #diazhendropriyono <a href="https://t.co/sxj2JBdEAE">https://t.co/sxj2JBdEAE</a>
Demokrat	17 April jgn lupa ke TPS dan Jeryas <b>Suara</b> Warba Kuning utk DPR RI Pusat, Mohon Dukungan dan Doanya PDemokrat 14 Calegnya Coblos No.5...Salam

Tweet pertama dibuat oleh perwakilan caleg dari PKPI, yang pada *tweetnya* terdapat kata ‘kampanye’, dan pada *tweet* kedua yang dibuat oleh caleg dari Partai Demokrat. Tweet yang kedua mempunyai kata ‘suara’, yang juga menjadi salah satu kata penyusun pada topik ‘Kampanye Politik’.

## BAB VII KESIMPULAN DAN SARAN

Pada bab ini akan dibahas mengenai kesimpulan penelitian permodelan topik yang telah dilakukan dan saran yang diusulkan untuk penelitian selanjutnya

### 7.1 Kesimpulan

Dari proses permodelan topik yang sudah dilakukan pada penelitian ini, dapat disimpulkan menjadi beberapa hal, antara lain:

1. Akuisisi data dilakukan dengan menggunakan Tweepy. Dari pengumpulan data *tweet* yang dilakukan sejak 22 Mei 2018 hingga 22 Mei 2019, didapatkan total 132.516 *tweet* yang dikumpulkan dari akun Twitter caleg dari 14 partai peserta Pemilu 2019.
2. Pengolahan data dimulai dengan melakukan pre-processing pada *tweet* yang sudah didapatkan. *Tweet* tersebut kemudian diolah menggunakan bantuan Gensim untuk melakukan permodelan topik menggunakan ATM dan LDA. Proses pembuatan model topik dengan menggunakan parameter *passes* sebanyak 1000 diketahui dapat menghasilkan beberapa model topik yang stabil, baik yang dibuat menggunakan ATM (*author* berupa partai) ataupun menggunakan LDA (kubu capres dan cawapres).
3. Hasil model yang didapatkan dari permodelan topik yang sudah dilakukan menggunakan ATM dan LDA kemudian divisualisasikan menggunakan Tableau. Visualisasi dilakukan dengan membuat 3 workbook, yaitu workbook rekap *tweet* Pemilu 2019, workbook hasil permodelan topik, dan workbook word cloud. Ketiga workbook tersebut dapat diakses secara online melalui Tableau Public.

## 7.2 Saran

Saran yang dapat diberikan untuk penelitian selanjutnya antara lain:

1. Perlu dilakukan pre-processing yang lebih detail pada *tweet*, terutama pada tahapan penghapusan *stopwords*. Dengan penghapusan *stopwords* yang lebih detail, tentunya akan mempengaruhi model topik yang dihasilkan.
2. Pengolahan data dengan menggunakan n-gram, untuk mengetahui model topik berdasarkan kata-kata yang berdekatan dan berkaitan.
3. Perlu dilakukan penghitungan dengan metode lain selain menggunakan *perplexity* dalam hal penentuan model dengan jumlah topik tertentu agar topik yang dihasilkan memiliki pendasaran dan perhitungan yang lebih rinci.
4. Pada penelitian selanjutnya dapat diimplementasikan juga pelabelan untuk mengetahui kata-kata yang dihasilkan pada *tweet* termasuk pada kategori positif atau negatif dan mengetahui tingkat kecenderungan topik tersebut.
5. Kedepannya dapat dibuat aplikasi permodelan topik *tweet* yang dapat secara otomatis melakukan *crawling* data pada *tweet* dan melakukan permodelan topik beserta visualisasi hasil model topik yang didapatkan.



## DAFTAR PUSTAKA

- [1] “KPU - Portal Publikasi Pemilihan Umum 2019.” [Online]. Available: <https://infopemilu.kpu.go.id/>. [Accessed: 07-Feb-2019].
- [2] *UNDANG-UNDANG REPUBLIK INDONESIA NOMOR 7 TAHUN 2017 TENTANG PEMILIHAN UMUM*. 2017.
- [3] G. Enli, “Twitter as arena for the authentic outsider: exploring the social media campaigns of Trump and Clinton in the 2016 US presidential election,” *Eur. J. Commun.*, vol. 32, no. 1, pp. 50–61, 2017.
- [4] S. Boulianne, “Social media use and participation: a meta-analysis of current research,” *Inf. Commun. Soc.*, vol. 18, no. 5, pp. 524–538, 2015.
- [5] E. W. T. Ngai, S. S. C. Tao, and K. K. L. Moon, “Social media research: Theories, constructs, and conceptual frameworks,” *Int. J. Inf. Manage.*, vol. 35, no. 1, pp. 33–44, 2015.
- [6] “Digital in 2018: World’s internet users pass the 4 billion mark - We Are Social.” [Online]. Available: <https://wearesocial.com/blog/2018/01/global-digital-report-2018>. [Accessed: 06-Feb-2019].
- [7] I. Sagarzazu and H. Klüver, “Coalition Governments and Party Competition: Political Communication Strategies of Coalition Parties,” *Polit. Sci. Res. Methods*, vol. 5, no. 02, pp. 333–349, 2017.
- [8] F. N. F. Insani, “Rancang Bangun Perangkat Lunak Teenstagram Untuk Mengelompokkan Topik Caption Akun Instagram Siswa Sekolah Menengah Pertama Di Surabaya,” 2017.
- [9] J. C. Na, *Digital Libraries: Providing Quality Information*, vol. 9469. 2015.

- [10] S. Jiang, X. Qian, J. Shen, Y. Fu, and T. Mei, "Author topic model-based collaborative filtering for personalized POI recommendations," *IEEE Trans. Multimed.*, vol. 17, no. 6, pp. 907–918, 2015.
- [11] D. Yu, J. Sun, Y. Wu, Z. Ni, and Y. Li, "Discovering Hidden Interests from Twitter for Multidimensional Analysis," pp. 329–334, 2017.
- [12] B. Mcnair, "An introduction to political communication," *Routledge*, p. 250, 2015.
- [13] "Ada 16 Parpol Nasional Peserta Pemilu 2019, Tahu Apa Saja? - kumparan.com." [Online]. Available: <https://kumparan.com/@kumparannews/ada-16-parpol-nasional-peserta-pemilu-2019-tahu-apa-saja>. [Accessed: 11-Feb-2019].
- [14] "Survei: Demokrat, PAN, Perindo Paling Tak Solid Dukung Capres." [Online]. Available: <https://www.cnnindonesia.com/nasional/20180927203752-32-333842/survei-demokrat-pan-perindo-paling-tak-solid-dukung-capres>. [Accessed: 25-Feb-2019].
- [15] "INFOGRAFIK: Bakal Caleg 16 Parpol pada Pemilu 2019 - Kompas.com." [Online]. Available: <https://nasional.kompas.com/read/2018/07/19/16200761/infografik-bakal-caleg-16-parpol-pada-pemilu-2019>. [Accessed: 11-Feb-2019].
- [16] J. H. Kietzmann, K. Hermkens, I. P. McCarthy, and B. S. Silvestre, "Social Media? Get serious! Understanding the functional building blocks of social media," vol. 82, no. 6, pp. 241–251, 2010.
- [17] R. Dolan, J. Conduit, J. Fahy, and S. Goodman, "Social media engagement behaviour: a uses and gratifications perspective," *J. Strateg. Mark.*, vol. 24, no. 3–4, pp. 261–277, 2016.
- [18] A. Heydon and M. Najork, "Mercator: A scalable, extensible Web crawler," vol. 2, pp. 1–15, 2011.

- [19] D. Blei, “Probabilistic Topic Models,” *Commun. ACM*, vol. 55, pp. 77–84, 2012.
- [20] D. Blei, M. Jordan, and A. Y. Ng, “Latent Dirichlet allocation,” *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, 2003.
- [21] M. Rosen-zvi, U. C. Irvine, and U. C. Irvine, “The Author-Topic Model for Authors and Documents,” *Mimno.Infosci.Cornell.Edu*, 2003.
- [22] “What is Tableau? Uses and Applications.” [Online]. Available: <https://www.guru99.com/what-is-tableau.html>. [Accessed: 02-Jul-2019].

*Halaman sengaja dikosongkan.*

**LAMPIRAN A. Hasil Pengujian Model menggunakan ATM dan LDA**

<http://bit.ly/2XnoNEN>

**LAMPIRAN B. Link Profil Tableau Public Visualisasi Permodelan Topik**

<https://public.tableau.com/profile/umi.shoima#!/>

**LAMPIRAN C. List Stopwords**

<https://github.com/umisz/atm/blob/master/stopwords.txt>

*Halaman sengaja dikosongkan.*

## BIODATA PENULIS



Penulis lahir di Surabaya pada tanggal 3 Januari 1998. Penulis merupakan anak pertama dari 2 bersaudara. Pendidikan formal yang ditempuh oleh penulis yaitu SD Hasyim Asy Ari, SMP Negeri 10 Surabaya, dan SMA Negeri 15 Surabaya.

Pada tahun 2015, penulis melanjutkan pendidikan ke jenjang S1 melalui jalur SNMPTN di Institut Teknologi Sepuluh Nopember (ITS) Surabaya yaitu di Departemen Sistem Informasi, Fakultas Teknologi Informasi dan Komunikasi (FTIK), terdaftar dengan NRP 05211540000047. Selama masa perkuliahan, penulis aktif mengikuti beberapa kegiatan kemahasiswaan, antara lain sebagai Sekretaris Departemen Entrepreneurship Himpunan Mahasiswa Sistem Informasi (HMSI) pada kepengurusan 2017/2018. Selain itu, dalam hal akademik penulis juga menjadi asisten dosen pada mata kuliah Desain Basis Data dan Sistem Enterprise.

Pada tahun ke-4 perkuliahan, penulis tertarik pada bidang *Social Media Analysis* dan mengambil bidang minat laboratorium Akuisisi Data dan Diseminasi Informasi (ADDI) dan menyelesaikan masa studi selama 8 semester. Penulis dapat dihubungi melalui email: [umishoima@gmail.com](mailto:umishoima@gmail.com).