

**SISTEM DETEKSI TOPIK POLITIK PADA TWITTER
MENGUNAKAN ALGORITMA *LATENT DIRICHLET*
*ALLOCATION***

SKRIPSI



KHAIRUL HUDHA NASUTION

5235154528

Skripsi ini Ditulis untuk Memenuhi Sebagian Persyaratan dalam Memperoleh
Gelar Sarjana

**PROGRAM STUDI PENDIDIKAN TEKNIK INFORMATIKA
DAN KOMPUTER**

FAKULTAS TEKNIK



UNIVERSITAS NEGERI JAKARTA

2019


**SISTEM DETEKSI TOPIK POLITIK PADA TWITTER MENGGUNAKAN
ALGORITMA LATENT DIRICHLET ALLOCATION**

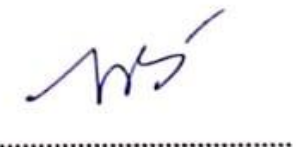
Nama : Khairul Hudha Nasution, NIM : 5235154528


HALAMAN PENGESAHAN

NAMA DOSEN	TANDA TANGAN	TANGGAL
Widodo, M.Kom. (Dosen Pembimbing 1)		20-08-19
Bambang Prasetya Adhi, S.Pd., M.Kom (Dosen Pembimbing 2)		20-8-19

PENGESAHAN PANITIA UJIAN SKRIPSI

NAMA DOSEN	TANDA TANGAN	TANGGAL
Diat Nurhidayat, M.Ti. (Ketua Penguji)		20-08-19

NAMA DOSEN	TANDA TANGAN	TANGGAL
Drs. Bachren Zaini, M.Pd. (Dosen Penguji 1)		19-8-19

NAMA DOSEN	TANDA TANGAN	TANGGAL
ZE. Ferdi Fauzan P., S.Pd., M.Pd.T. (Dosen Penguji 2)		19 Agus 2019

Tanggal Lulus : 29 Agustus 2019

HALAMAN PERNYATAAN

Dengan ini peneliti menyatakan bahwa :

1. Karya tulis skripsi peneliti yang berjudul Sistem Deteksi Topik Politik Pada Twitter Menggunakan Algoritma *Latent Dirichlet Allocation* adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik sarjana, baik di Universitas Negeri Jakarta maupun di perguruan tinggi lainnya;
2. Karya tulis skripsi yang berjudul Sistem Deteksi Topik Politik Pada Twitter Menggunakan Algoritma *Latent Dirichlet Allocation* adalah murni gagasan, rumusan, dan penelitian peneliti sendiri dengan arahan dari dosen pembimbing;
3. Dalam karya tulis, tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasi orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan dicantumkan dalam daftar pustaka;
4. Pernyataan peneliti dibuat dengan sesungguhnya dan apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka peneliti bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh karena karya tulis ini, serta sanksi lainnya sesuai norma yang berlaku di Universitas Negeri Jakarta.

Jakarta, 19 Agustus 2019

Yang Membuat Pernyataan



Khairul Hudha Nasution

5235154528

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Allah SWT yang telah memberikan rahmat serta nikmat-Nya dan tak lupa shalawat serta salam senantiasa dicurahkan kepada Nabi Muhammad SAW yang telah menunjukkan jalan ke zaman yang terang benderang. Berdasarkan hal tersebut penulis dapat menyelesaikan skripsi penulis dengan judul “Sistem Deteksi Topik Politik Pada Twitter Menggunakan Algoritma *Latent Dirichlet Allocation*” yang merupakan persyaratan untuk meraih gelar Sarjana Pendidikan.

Skripsi ini dapat terwujud dengan baik berkat bimbingan, dorongan, serta saran-saran, dan bantuan dari berbagai pihak. Pada kesempatan ini, penulis ingin menyampaikan ucapan terimakasih yang sebesar-besarnya kepada:

1. Ibu Yuliatr Sastrawijaya, M.Pd. selaku Ketua Program Studi Pendidikan Teknik Informatika dan Komputer Fakultas Teknik Universitas Negeri Jakarta;
2. Bapak Widodo, M.Kom. selaku dosen pembimbing I dan Bapak Bambang Prasetya Adhi, S.Pd., M.Kom selaku dosen pembimbing II yang telah memberikan motivasi, arahan, dan kepercayaan kepada penulis dalam menyelesaikan skripsi ini;
3. Kedua orang tua, Bapak Klolad Nasution dan Ibu Kholilah Pulungan serta adik Khairul Ikhsan Nasution, dan Kaela Khadijah yang selalu memberi semangat, mendo'akan, dan memotivasi penulis dalam menyelesaikan skripsi ini;
4. Azizah Khoiro Nisah yang membantu penulis dalam menyelesaikan skripsi ini;
5. Teman-teman PTIK 2015 UNJ yang senantiasa memberikan saran, pengalaman, semangat, dan do'a dalam penyusunan skripsi ini;
6. Seluruh pihak yang terlibat secara langsung maupun tidak langsung yang tidak bisa disebutkan satu per satu.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan, maka dari itu penulis memohon maaf apabila terdapat kekurangan dan kesalahan, baik dari isi maupun tulisan. Akhir kata penulis berharap agar penulisan dan

penyusunan skripsi ini dapat bermanfaat bagi para pembaca dan semua pihak terkait.

Jakarta, 19 Agustus 2019

Khairul Hudha Nasution

SISTEM DETEKSI TOPIK PADA TWITTER MENGGUNAKAN ALGORITMA *LATENT DIRICHLET ALLOCATION*

KHAIRUL HUDHA NASUTION

ABSTRAK

Tahun 2019 Indonesia melaksanakan tahun politik, banyak terjadi peristiwa politik yang membuat masyarakat Indonesia menyikapi dengan berbagai macam tanggapan dari berbagai banyak tanggapan tersebut beberapa dituliskan dalam media sosial Twitter dan data tersebut dapat diolah untuk menggambarkan bagaimana pendapat masyarakat akan suatu kejadian politik. Penelitian ini bertujuan untuk mendapatkan analisis dari implementasi algoritma LDA untuk menentukan topik politik pada Twitter. Metode yang digunakan adalah dengan algoritma LDA yang digunakan untuk menghitung kemungkinan topik yang ada untuk setiap *tweet*-nya, LDA merupakan model probabilistik yang dapat menggambarkan topik tanpa perlu melakukan proses klasifikasi sebelumnya, sistem akan otomatis mendeteksi topik-topik yang ada. Hasil penelitian dengan pengujian 3 kali dengan jumlah data masing-masing 100, 1000 dan 6000 dengan menggunakan setingan LDA bawaan dari *library* Genism dan jumlah topik 10 menghasilkan rata-rata nilai kebenaran 90%. Sehingga dapat disimpulkan bahwa LDA dapat digunakan dan memiliki nilai kebenaran yang tinggi dalam mendeteksi topik politik pada Twitter

Kata kunci: *Latent Dirichlet Allocation, Natural Leanguage Processing, Twitter*

LATENT DIRICHLET ALLOCATION ALGORITHM FOR TOPIC DETECTION SYSTEM ON TWITTER.

KHAIRUL HUDHA NASUTION

ABSTRACT

In 2019 Indonesia held a political year, many political events took place that made the Indonesian people respond with a variety of responses from various responses many written in social media Twitter and the data can be processed to illustrate how the public opinion of a political event. This study goals to obtain an analysis of the implementation of the LDA algorithm to determine political topics on Twitter. The method used is the LDA algorithm that is used to calculate the possible topics for each tweet, LDA is a probabilistic model that can describe topics without the need to prior classification process, the system will automatically detect existing topics. The results of the study by testing 3 times with the amount of data respectively 100, 1000 and 6000 using the default LDA settings from the Genism library and the number of topics 10 produced an average truth value of 90%. So it can be concluded that LDA can be used and has a high truth value in detecting political topics on Twitter

Keywords: Latent Dirichlet Allocation, Natural Leanguage Processing, Twitter

DAFTAR ISI

KATA PENGANTAR.....	iii
DAFTAR ISI.....	viii
DAFTAR TABEL	x
DAFTAR GAMBAR.....	xi
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Identifikasi Masalah	3
1.3. Batasan Masalah	4
1.4. Rumusan Masalah	4
1.5. Tujuan Penelitian	4
1.6. Manfaat Penelitian	4
BAB II TINJAUAN PUSTAKA	5
2.1. Kerangka Teoritik	5
2.1.1. Deteksi Topik	5
2.1.2. <i>Latent Dirichlet Allocation</i>	5
2.1.3. Politik.....	7
2.1.4. Twitter.....	8
2.2. Penelitian Relevan.....	8
2.3. Kerangka Berpikir	13
BAB III METODE PENELITIAN.....	15
3.1. Tempat dan Waktu Penelitian.....	15
3.2. Alat dan Bahan Penelitian	15
3.2.1. Alat Penelitian	15
3.2.2. Bahan Penelitian	16
3.3. Diagram Alir Penelitian.....	17
3.3.1. Pengambilan Data.....	17
3.3.2. Preprocessing.....	18
3.3.3. Penerapan LDA	20
BAB IV HASIL PENELITIAN	25
4.1. Deskripsi Hasil Penelitian	25
4.2. Hasil Penelitian	25

4.2.1.	Pengambilan Data	25
4.2.2.	<i>PreProcesing</i>	27
4.2.3.	<i>Latent Dirichlet Allocation (LDA)</i>	29
4.2.4.	Visualisasi.....	32
4.2.5.	Evaluasi	34
4.3.	Pembahasan	34
BAB V PENUTUP		36
5.1.	Kesimpulan	36
5.2.	Saran.....	36
DAFTAR PUSTAKA.....		38
LAMPIRAN-LAMPIRAN.....		39
DAFTAR RIWAYAT HIDUP.....		50

Daftar Tabel

Tabel 2.1 Penelitian Relevan.....	13
Tabel 3.1 Contoh Penerapan Sebelum dan Sesudah <i>Lowercase</i>	19
Tabel 3.2 Contoh Penerapan Sebelum dan Sesudah Tokenisasi	20
Tabel 3.3 Contoh Penerapan Sebelum dan Sesudah <i>Stopwords</i>	20
Tabel 3.4 Contoh Penerapan Sebelum dan Sesudah <i>Stemming</i>	20
Tabel 3.5 Keterangan Notasi dan Definisi Persamaan	22
Tabel 4.1 Perbandingan Waktu <i>Preprocessing</i> per Jumlah Data	28
Tabel 4.2 Perbandingan Waktu Pemodelan LDA	31
Tabel 4.3 Evaluasi LDA	34

Daftar Gambar

Gambar 2.1 Visualisasi LDA	7
Gambar 2.2 Kerangka Berpikir	15
Gambar 3.1 Diagram Alir Penelitian	17
Gambar 3.2 Ilustrasi Cara Kerja LDA	23
Gambar 3.3 Ilustrasi Hasil Matplotlib dengan 4 Topik	24
Gambar 4.1 Data Sebelum dan Sesudah <i>Filtering</i>	26
Gambar 4.2 Proses <i>Lowercase</i>	27
Gambar 4.3 Proses Tokenisasi	27
Gambar 4.4 Proses <i>Stopwords</i>	27
Gambar 4.5 Proses <i>Stemming</i>	28
Gambar 4.6 Kamus yang Digunakan LDA	29
Gambar 4.7 <i>Corpus</i> yang Digunakan LDA	30
Gambar 4.8 Model LDA	31
Gambar 4.9 Visualisasi 100 <i>Tweet</i>	32
Gambar 4.10 Visualisasi 1000 <i>Tweet</i>	33
Gambar 4.11 Visualisasi 6000 <i>Tweet</i>	33

BAB I

PENDAHULUAN

1.1. Latar Belakang

Informasi merupakan data yang telah diberi makna melalui konteks. Informasi sendiri dapat disampaikan dengan berbagai macam cara dan media, mulai dari bentuk gambar, suara ataupun tulisan, tergantung dari tujuan dan keinginan penyampai atau pembuat untuk menyampaikan informasi tersebut, dan informasi dalam bentuk tulisan adalah model penyampaian informasi yang paling banyak digunakan karena kemudahan dalam membuat, menyebarkan dan banyaknya informasi yang disampaikan dalam satu tulisan.

Ada 2 macam teks, yaitu teks panjang dan teks pendek. Teks panjang biasa kita temui pada paragraf, dimana paragraf memiliki banyak kata dan teks pendek yang biasa kita temui pada kalimat atau judul, dimana kalimat tidak memiliki banyak kata di dalamnya, sehingga memungkinkan penyampain informasi yang tidak sesuai. Dalam menentukan isi kandungan dari suatu berita atau paragraf kita bisa melakukan dengan menghitung kata unik yang paling banyak keluar, sedangkan pada teks pendek hal tersebut cenderung lebih sulit dilakukan karena keterbatasan kata yang ada dalam teks pendek.

Pada tahun ini (2019) Indonesia menjalani pesta politik yaitu pemilihan wakil rakyat dan presiden dan wakilnya yang dilaksanakan 5 tahun sekali, reaksi masyarakatpun bermacam-macam dalam mengomentari terhadap tindakan dan peristiwa politik yang ada. Salah satu komentar-komentar yang disampaikan tertulis di media sosial dimana masyarakat dapat dengan bebas menyampaikan pendapatnya atas kejadian politik yang terjadi. Dari media sosial pula dapat

menjadi tolak ukur tanggapan masyarakat atas suatu putusan atau suatu kejadian tertentu baik buruknya suatu kejadian di pandangan masyarakat umum, bahkan dapat menggambarkan secara tidak langsung menang kalahnya pasangan calon berdasarkan jumlah dan sentimen publik pada beberapa tokoh.

Salah satu media sosial yang paling populer adalah Twitter. Menurut (Yudha Pratomo, 2019) yang dikutip dari tekno.kompas.com pada kuartal empat 2018 tercatat rata-rata *user* aktif di Indonesia perharinya kurang lebih 126 juta *user*. Pengguna Twitter dapat menulis (*tweet*) tulisan hingga batas maksimal 280 karakter.

Untuk mengetahui tanggapan orang pada media sosial perlu dikumpulkannya data dan melakukan penggambaran berdasarkan tanggapan masyarakat, proses ini biasanya memakan banyak waktu karena memproses banyak data, dan proses yang dilakukan adalah proses yang repetitif yang artinya dapat digantikan dengan komputer. Dalam menyelesaikan masalah tersebut ilmu teknologi sekarang ini memiliki dua solusi utama yaitu menggunakan *machine learning* atau *artificial intelligence*. *Artificial Intelligence* adalah sistem yang bertujuan untuk melakukan apa yang dilakukan manusia dengan lebih baik sedangkan *machine learning* bertujuan untuk belajar dan bertindak berdasarkan apa yang dipelajari, pada kesempatan kali ini peneliti memilih menggunakan *machine learning* karena dirasa sudah cukup dan paling baik digunakan pada kasus ini.

Dalam *machine learning* terdapat *natural language processing* yang intinya bekerja dengan mencoba memahami tulisan manusia, untuk memahami tulisan manusia ada teknik yang namanya *topic modeling* atau permodelan topik yang bekerja untuk mencari inti topik dari kumpulan-kumpulan kata atau dokumen.

Didalam permodelan topik ada dua cara populer yaitu *Latent Dirichlet Allocation* dan *Latent Semantic Analysis* atau *Probabilistic Latent Semantic Analysis*.

LDA (*Latent Dirichlet Allocation*) merupakan model probabilistik generatif dari koleksi data diskrit, dimana LDA akan memberikan hasil perhitungan antara *document*, *topics* dan *words* untuk menentukan tingkat kesesuaian antar *document* (pada kasus ini *tweet*). LDA sendiri dipilih sebagai algoritma yang digunakan karena dirasa paling cocok dan paling tinggi tingkat ke-akuratannya untuk menggambarkan *topic* dari dokumen dibanding model lainnya dan sebagai pengujian seberapa bagus implementasinya pada teks pendek (Twitter).

Kita sebagai manusia terkadang mencari sesuatu berdasarkan topik yang kita senangi, tak terkecuali pada tulisan yang kita baca pada media sosial Twitter, kita sebagai manusia mungkin mengetahui dan bisa memilih mana yang sesuai dengan yang kita cari dan butuhkan mana yang tidak, namun komputer tidak memahami apa yang ditulis manusia. Oleh karena itu diharapkan dengan menerapkan implementasi LDA untuk menentukan topik dapat membuat komputer dapat menggambarkan topik pada *tweet* khususnya pada kasus politik dan umumnya pada teks pendek.

1.2. Identifikasi Masalah

Berdasarkan latar belakang yang telah dijelaskan, identifikasi masalah pada penelitian ini adalah :

1. Dengan menggunakan metode manual dibutuhkan waktu yang lama dalam penggambaran sentimen dari media sosial.

2. Belum diketahuinya seberapa bagus implementasi algoritma *Latent Dirichlet Allocation* pada teks pendek khususnya kasus penentuan topik politik Indonesia pada Twitter.

1.3. Batasan Masalah

Melihat luasnya lingkup permasalahan, agar memperoleh hasil penelitian yang optimal dan terfokus, maka penelitian dibatasi pada pembuatan aplikasi atau *prototype* untuk menentukan topik. Batasan masalah terdiri dari:

1. Data yang diambil dari Twitter yang berbahasa Indonesia.
2. Pengujian hanya menggunakan algoritma *Latent Dirichlet Allocation* saja dan tidak dibandingkan dengan algoritma lainnya .

1.4. Rumusan Masalah

Berdasarkan pembatasan masalah, maka rumusan masalah yang akan dibahas pada penelitian ini adalah, bagaimana analisis dan implementasi algoritma *Latent Dirichlet Allocation* pada penentuan topik politik pada Twitter?.

1.5. Tujuan Penelitian

Tujuan Penelitian ini adalah untuk menganalisis dan mengimplementasikan kegunaan algoritma *Latent Dirichlet Allocation* dalam penentuan topik dari tulisan teks pendek dalam media Twitter.

1.6. Manfaat Penelitian

Manfaat dari penelitian ini adalah agar:

1. Mengetahui cara kerja atau metode dan tingkat ke-akuratan implementasi algoritma *Latent Dirichlet Allocation* pada penentuan topik pada teks pendek.
2. Sebagai bahan kajian untuk penelitian berikutnya mengenai teks pendek dan *Latent Dirichlet Allocation*.

BAB II

TINJAUAN PUSTAKA

2.1. Kerangka Teoritik

Berikut teori teori yang dirasa mendukung penelitian sehingga penelitian ini dibuat.

2.1.1. Deteksi Topik

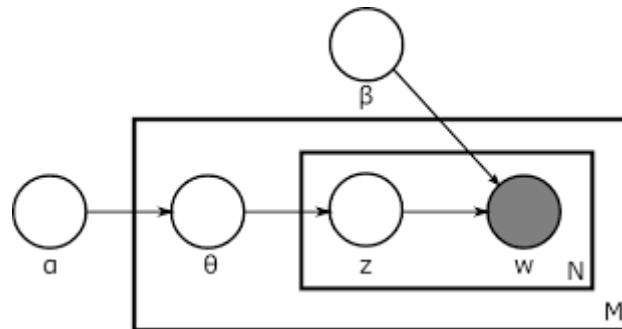
Menurut (James Allan, 2002) tugas deteksi topik adalah dapat menentukan topik dari suatu novel yang sebelumnya belum diketahui. Dalam pengerjaannya topik ditentukan dari cerita yang dibahasnya, namun sistem tidak diberi tahu topiknya. Sehingga sistem harus memiliki kemampuan untuk mengetahui dan memahami topik dari cerita, dan kinerja sistem haruslah *independent*.

Dalam KBBI sendiri, deteksi artinya usaha menemukan dan menentukan keberadaan, anggapan, atau kenyataan. Topik artinya pokok pembicaraan dalam diskusi, ceramah, karangan, dan sebagainya; bahan diskusi. Berdasarkan penjelasan diatas dapat ditarik kesimpulan bahwa deteksi topik adalah kegiatan untuk menentukan pokok pembicaraan dalam suatu diskusi atau tulisan.

2.1.2. Latent Dirichlet Allocation

Menurut (Blei, 2003) *Latent Dirichlet Allocation* (LDA) merupakan model probabilistik generative dari kumpulan tulisan yang disebut *corpus*. Ide dasarnya adalah setiap dokumen adalah hasil presentasi berbagai macam topik acak, yang mana setiap topik memiliki karakter yang ditentukan berdasarkan distribusi kata-kata yang terdapat di dalamnya. Menurut (Campbell, 2014) LDA dapat digunakan untuk meringkas, melakukan klasterisasi, menghubungkan maupun memproses

data yang sangat besar karena LDA menghasilkan daftar topik yang diberi bobot untuk masing-masing dokumen.



Gambar 2.1. Visualisasi LDA

Blei memvisualisasikan LDA seperti gambar 2.1, dimana terdapat tiga tingkatan pada representasi LDA. Variable α dan β sebagai variable tingkat *corpus*, variable θ sebagai variable tingkat dokumen (M) dan variabel Z dan W sebagai variable tingkat kata(N). Dengan variable α sebagai banyak distribusi topik pada dokumen, semakin besar nilainya menandakan campuran topik yang dibahas di dalam dokumen semakin banyak. Variable β sebagai banyak distribusi kata dalam topik, semakin tinggi nilainya maka semakin banyak kata di dalam topik. Variable θ sebagai distribusi topik untuk dokumen tertentu, semakin tinggi nilainya maka semakin banyak topik dalam satu dokumen. Variable Z mempersentasikan topik dari kata tertentu pada sebuah dokumen dan variable W mempersentasikan kata yang berkaitan dengan topik tertentu yang terdapat di dalam dokumen. Bentuk lingkaran mempersentasikan individual kata, lingkaran abu-abu yang diteliti dan yang kosong yang tidak secara langsung diteliti, jika kita mengambil kemungkinan marginal dari satu dokumen, maka dapat rumusan dari LDA dapat didefinisi sebagai berikut:

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

Menurut (Campbell, 2014) secara umum LDA bekerja dengan memberi *input* beberapa parameter yang menghasilkan keluaran berupa model yang memiliki bobot sehingga model tersebut dapat dinormalisasi sesuai probabilitas. Probabilitas mengacu pada dua jenis yakni (a) probabilitas bahwa suatu dokumen spesifik akan menghasilkan topik yang spesifik pula dan (b) probabilitas bahwa suatu topik spesifik menghasilkan kata-kata spesifik tertentu. Probabilitas jenis (a), dokumen yang sudah diberi label dengan daftar topik seringkali dilanjutkan hingga menghasilkan probabilitas jenis (b), yang menghasilkan kata-kata spesifik tertentu.

2.1.3. Politik

Menurut (Miriam Budiardjo, 2007) dalam bukunya mengatakan politik dalam suatu negara (*state*) berkaitan dengan masalah kekuasaan (*power*) pengambilan keputusan (*decision making*), kebijakan publik (*public policy*), dan alokasi atau distribusi (*allocation or distribution*). Dewasa ini definisi mengenai politik yang sangat normatif itu telah terdesak oleh definisi-definisi lain yang lebih menekankan pada upaya (*means*) untuk mencapai masyarakat yang baik, seperti kekuasaan, pembuatan keputusan, kebijakan, alokasi nilai, dan sebagainya.

Pada umumnya dapat dikatakan bahwa politik (*politics*) adalah usaha untuk menentukan peraturan-peraturan yang dapat diterima baik oleh sebagian besar warga, untuk membawa masyarakat ke arah kehidupan bersama yang harmonis. Usaha menggapai *the good life* ini menyangkut bermacam-macam kegiatan yang antara lain menyangkut proses penentuan tujuan dari sistem, serta cara-cara

melaksanakan tujuan itu. Masyarakat mengambil keputusan mengenai apakah yang menjadi tujuan dari sistem politik itu dan hal ini menyangkut pilihan antara beberapa alternatif serta urutan prioritas dari tujuan-tujuan yang telah ditentukan itu.

2.1.4. Twitter

Twitter adalah berita *online* dan jejaring sosial dimana orang berkomunikasi dengan teks pendek yang disebut *tweet*. *Tweet* yang di *publish* akan dikirimkan kepada orang-orang yang mengikutinya di Twitter. Twitter juga dapat dideskripsikan sebagai *microbloging* (Paul Gil, 2019).

Twitter bekerja dengan cara mengirimkan dan menerima pesan pendek dari dan kepada yang mengikuti dan diikuti di Twitter. Twitter juga dapat mem*follow* dan *unfollow user*, *retweet* dan *hashtag*. Twitter dapat digunakan sebagai sosial media, layanan berita, alat bantu jualan, forum, layanan pesan dan lain sebagainya.

Pada bulan September dilakukan uji coba pengubahan batasan karakter dari yang awalnya 140 karakter per *tweet* menjadi 280 karakter, kecuali pada bahasa Jepang, Korea dan China karena perbedaan bahasa. Uji coba tersebut diterapkan secara global pada 7 November 2017 (Aliza Rosen, 2017).

2.2. Penelitian Relevan

1. Analisis Klasifikasi Opini *Tweet* pada Media Sosial Twitter Menggunakan *Latent Dirichlet Allocation* (LDA) (Minanti & Retno, 2019: 185)

Penelitian dilakukan oleh Minanti Fitriasih dan Retno Kusumaningrum di Departement Ilmu Komputer/Informatika Fakultas Sains dan Matematika, Universitas Diponegoro dalam acara Seminar Nasional Teknologi Informasi dan

Komunikasi 2019 (SENTIKA 2019) dan merupakan hasil penelitian jurnal tahun 2019.

Penelitian ini bertujuan untuk mengembangkan aplikasi yang dapat mengklasifikasikan opini *tweet* sebagai *neutral opinion* atau *crime indicated opinion*. Dari penelitian ini disimpulkan bahwa pengaruh jumlah topik terhadap nilai akurasi yaitu jumlah topik yang kecil memberikan nilai akurasi tertinggi namun kurang stabil dibandingkan dengan jumlah topik yang besar apabila diberikan perubahan parameter pada nilai *alfa* dan *beta*.

2. Sistem Monitoring Percakapan pada Toko *Online* Menggunakan Metode *Latent Dirichlet Allocation* (LDA) Studi Kasus: Toko *Online* “berrybenka.com” (Wisnu, 2018: 58)

Penelitian dilakukan oleh Wisnu Kurniawan di Fakultas Teknologi Industri Universitas Islam Indonesia dan merupakan hasil penelitian skripsi pada tahun 2018. Penelitian ini bertujuan untuk mengidentifikasi apakah metode LDA dapat dengan baik mengidentifikasi topik-topik yang sering dibicarakan antara *customer service* dengan pembeli pada toko *online* Berrybenka. Dari penelitian ini disimpulkan bahwa Metode LDA terbukti dapat melakukan pemodelan topik atau menampilkan kata-kata yang sering dibahas terhadap data percakapan antara *customer service* dan pembeli. Hal ini merupakan hal yang bermanfaat bagi bisnis perusahaan karena dapat mengetahui keadaan sebenarnya yang terjadi. Untuk melakukan pemodelan topik pada data percakapan, dilakukan beberapa langkah yaitu *preprocessing*, pembuatan model topik dengan LDA, dan penentuan cacah topik terbaik dengan menghitung nilai *topic coherence*. Pada penelitian ini terbukti bahwa *preprocessing* merupakan sesuatu yang sangat penting dalam

menghasilkan keluaran yang baik, terutama dalam proses menggabungkan kata yang terpisah namun satu makna dan mengubahnya menjadi satu kesatuan.

3. Analisis Topik Informasi Publik Media Sosial di Surabaya Menggunakan Pemodelan *Latent Dirichlet Allocation* (LDA) (I Made Kusnanta Bramantya Putra, 2017 :159)

Penelitian dilakukan oleh I Made Kusnanta Bramantya Putra di Fakultas Teknologi Informasi Institut Teknologi Sepuluh Nopember Surabaya dan merupakan hasil penelitian skripsi pada tahun 2017. Penelitian ini bertujuan untuk Melakukan analisis *topic modeling* untuk mengetahui topik-topik apa yang sedang dibahas pada media sosial informasi publik di Surabaya, dan untuk Melakukan validasi pada luaran dari *topic modeling* untuk mengukur kinerja metode LDA pada dokumen. Dari penelitian ini disimpulkan bahwa model yang dihasilkan dengan metode LDA pada studi kasus ini dapat diinterpretasi manusia dengan baik.

4. Implementasi Metode Latent Dirichlet Allocation (LDA) untuk Menentukan Topik Teks Berita. (Dimas Aryanto Saputro, 2019)

Penelitian dilakukan oleh Dimas Aryanto Saputro di Teknik Informatika Fakultas Teknis Universitas Nusantara PGRI Kediri dan merupakan hasil penelitian skripsi pada tahun 2019. Penelitian ini bertujuan untuk mempermudah perusahaan media cetak koran harian dalam pemilihan *headline* berita dengan cara otomatis (menggunakan metode LDA). Dari penelitian ini disimpulkan bahwa metode LDA dapat digunakan untuk menentukan topik berita dan memiliki hasil yang sama dengan perhitungan pada sistem yang menggunakan metode *Latent*

Dirichlet Allocation, maka dihasilkan urutan rekomendasi berita yang dijadikan *headline* sesuai dengan bobotnya.

5. Pemodelan Topik Untuk Media Sosial Menggunakan *Latent Dirichlet Allocation* Studi kasus: Analisis Tren Berita dalam Media Sosial. (Ruske Illa Kengken, 2014: 54)

Penelitian dilakukan oleh Ruske Illa Kengken di Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Gadjah Mada Yogyakarta dan merupakan hasil penelitian skripsi pada tahun 2014. Penelitian ini bertujuan untuk menerapkan model LDA dalam pemodelan topik untuk mencari topik utama dalam data tekstual yang besar, menerapkan metode *Bayesian* dalam estimasi parameter model LDA, dan mengaplikasikan model LDA dalam pemodelan topik untuk mendapatkan tren berita yang paling dibicarakan masyarakat di Twitter. Dari penelitian ini disimpulkan bahwa model ini dapat mengelompokkan kata-kata yang memiliki kemiripan pada setiap topik yang terbentuk sehingga dapat diinterpretasikan menjadi suatu kalimat yang lebih bermakna dan informatif bila dibandingkan dengan perhitungan frekuensi kemunculan kata, model LDA dibentuk melalui *generative process* dan dapat diestimasi menggunakan metode *Bayesian* yang estimasi parameter modelnya dengan menggunakan metode *Bayesian* dapat dilakukan dengan menggunakan *Gibbs Sampling*.

Tabel 2.1 Penelitian Relevan

No	Judul Penelitian	Publikasi	Metode	Hasil/Kesimpulan
1	Analisis Klasifikasi Opini <i>Tweet</i> pada Media Sosial Twitter Menggunakan <i>Latent Dirichlet Allocation</i> (LDA)	Jurnal-Seminar Nasional Teknologi Informasi dan Komunikasi 2019 (SENTIKA 2019)	<i>Latent Dirichlet Allocation</i> untuk mengetahui pengaruh jumlah topik terhadap nilai akurasi	Pengaruh jumlah topik terhadap nilai akurasi yaitu jumlah topik yang kecil memberikan nilai akurasi tertinggi namun kurang stabil dibandingkan dengan jumlah topik yang besar apabila diberikan perubahan parameter pada nilai <i>alfa</i> dan <i>beta</i>
2	Sistem Monitoring Percakapan pada Toko <i>Online</i> Menggunakan Metode <i>Latent Dirichlet Allocation</i> (LDA) Studi Kasus: Toko <i>Online</i> “berrybenka.com”	Skripsi-Fakultas Teknologi Industri Universitas Islam Indonesia	<i>Latent Dirichlet Allocation</i> dalam teks <i>chat</i>	Metode LDA terbukti dapat melakukan pemodelan topik atau menampilkan katakata yang sering dibahas terhadap data percakapan antara <i>customer service</i> dan pembeli
3	Analisis Topik Informasi Publik Media Sosial di Surabaya Menggunakan Pemodelan <i>Latent Dirichlet Allocation</i> (LDA)	Skripsi-Teknologi Informasi Institut Teknologi Sepuluh Nopember Surabaya	<i>Latent Dirichlet Allocation</i>	Model yang dihasilkan dengan metode LDA pada studi kasus ini dapat diinterpretasi manusia dengan baik
4	Implementasi Metode <i>Latent Dirichlet Allocation</i> (LDA) untuk Menentukan Topik Teks Berita	Skripsi-Fakultas Teknik Informatika Fakultas Teknis Universitas Nusantara PGRI Kediri	<i>Latent Dirichlet Allocation</i> dalam teks berita	Metode <i>latent dirichlet allocation</i> dapat digunakan untuk menentukan topik berita.

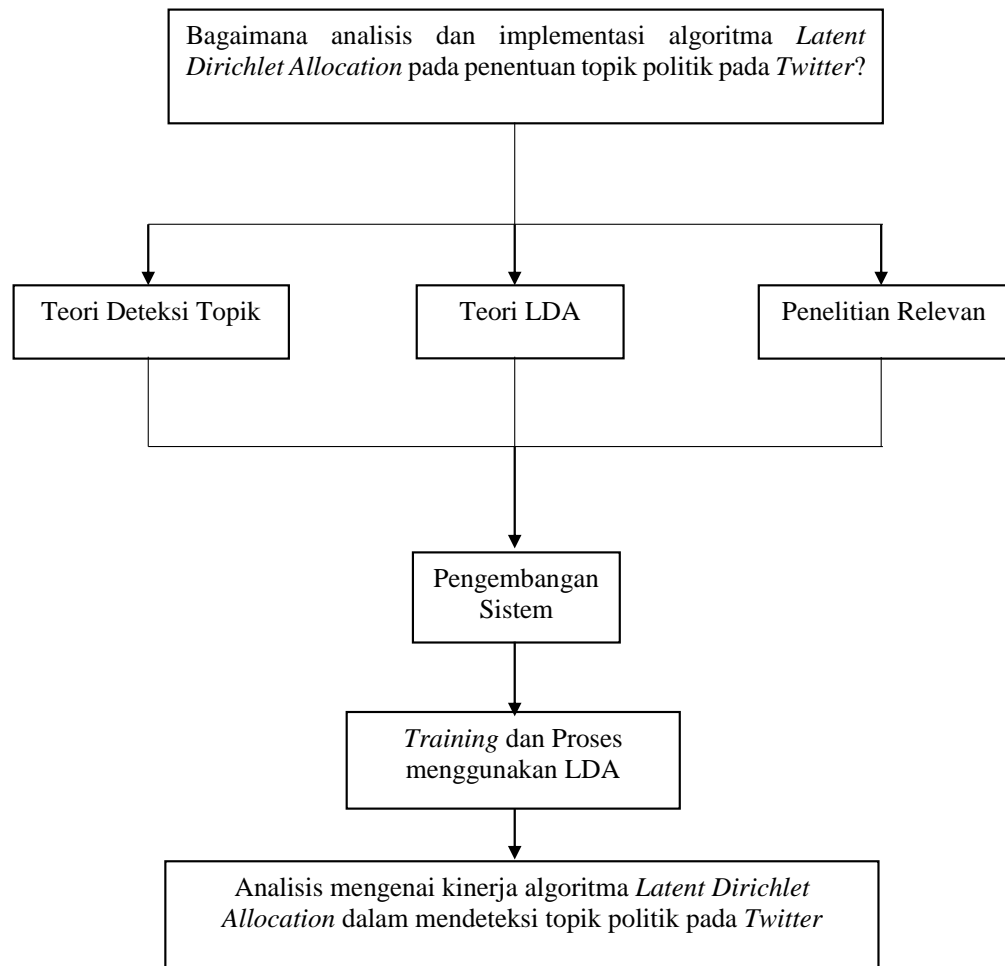
No	Judul Penelitian	Publikasi	Metode	Hasil/Kesimpulan
5	Pemodelan Topik Untuk Media Sosial Menggunakan <i>Latent Dirichlet Allocation</i> Studi kasus: Analisis Tren Berita dalam Media Sosial	Skripsi-Statistik Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Gadjah Mada Yogyakarta	<i>Latent Dirichlet Allocation</i> digabungkan dengan <i>Gibbs sampling</i>	Estimasi parameter dalam model <i>Latent Dirichlet Allocation</i> dengan metode Bayesian dapat dilakukan dengan menggunakan <i>Gibbs sampling</i> .

2.3. Kerangka Berpikir

Berdasarkan latar belakang masalah dan kajian pustaka, metode LDA dirasa dapat memberikan hasil penelitian yang spesifik. Ditambah lagi dengan dukungan dari penelitian terkait yang dilakukan oleh Wisnu Kurniawan di Fakultas Teknologi Industri Universitas Islam Indonesia pada tahun 2018 dengan judul “Sistem *Monitoring* Percakapan pada Toko *Online* Menggunakan Metode *Latent Dirichlet Allocation* (LDA) Studi Kasus: Toko *Online* berrybenka.com.” yang mendapat kesimpulan Metode LDA terbukti dapat melakukan pemodelan topik atau menampilkan kata-kata yang sering dibahas terhadap data percakapan antara *customer service* dan pembeli.

Dari beberapa penelitian tersebut peneliti tertarik untuk menggunakan metode LDA untuk menentukan topik politik pada media sosial Twitter. Untuk melakukan penelitian tersebut peneliti harus melakukan pembuatan sistem, setelah sistem dibuat maka dilakukan proses LDA dengan dua tahapan utama *training* dan proses pengolahan lalu melakukan optimalisasi agar tingkat akurasi yang tinggi pada teks pendek yang awalnya metodenya ditujukan untuk teks panjang.

Hasil dari pengujian akan menghasilkan analisis yang diharapkan dapat menjawab rumusan masalah.



Gambar 2.2. Diagram Alir

BAB III

METODE PENELITIAN

3.1. Tempat dan Waktu Penelitian

Penelitian ini dilakukan di Laboratorium Komputer Program Studi Pendidikan Teknik Informatika dan Komputer, Gedung Elektro (L1), Fakultas Teknik Universitas Negeri Jakarta (Kampus A) yang berlokasi di Jalan Rawamangun Muka, Jakarta Timur, 13220. Penelitian ini dilakukan sejak bulan Januari – Agustus 2019.

3.2. Alat dan Bahan Penelitian

Alat yang digunakan dalam penelitian ini merupakan komputer dan *library* yang nantinya akan digunakan untuk membuat prototipe dan menganalisis data, sedangkan data sendiri diambil dari tanggapan masyarakat pada *tweet*, berikut penjelasan lengkapnya.

3.2.1. Alat Penelitian

Alat-alat yang digunakan dalam melaksanakan penelitian ini, yaitu berupa perangkat keras, antara lain:

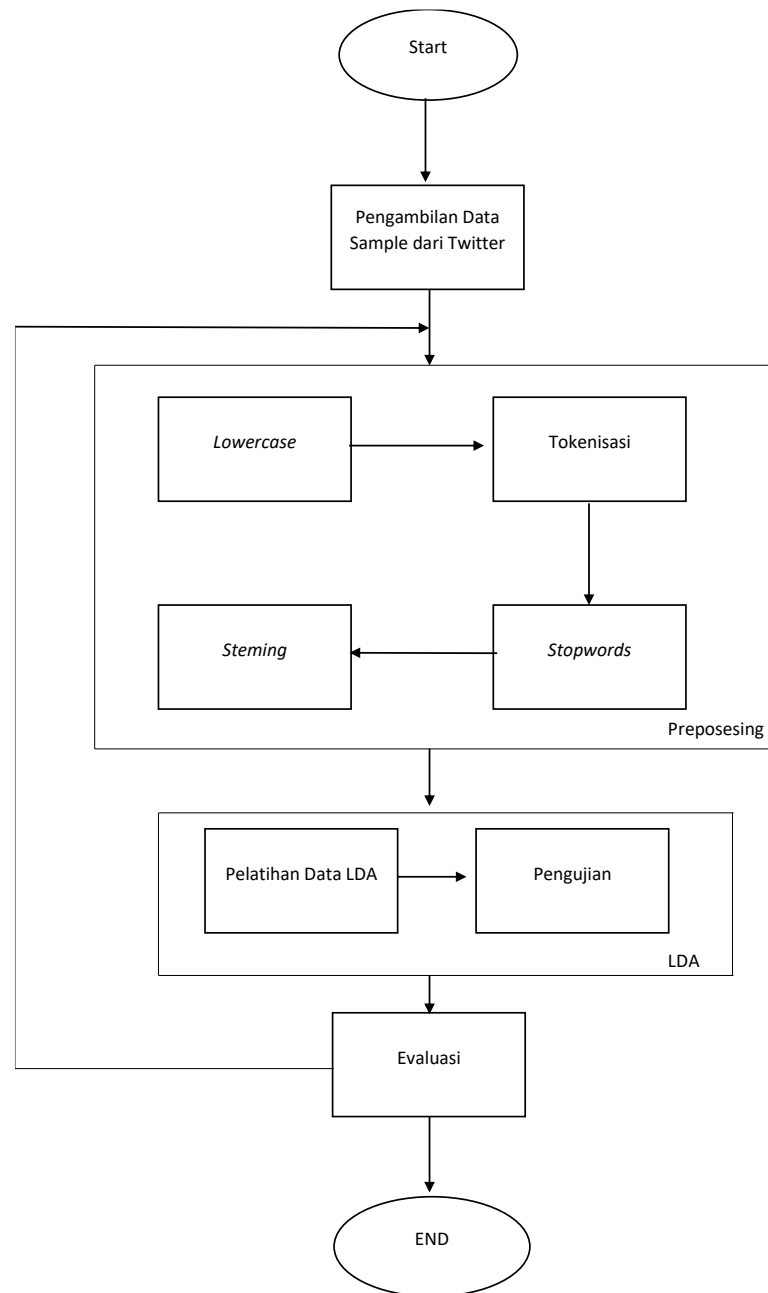
1. Komputer dengan *prosesor* AMD *Ryzen* 2400g dan RAM 8GB.
2. Python 3, sebagai bahasa pemrograman.
3. Pycharm, sebagai IDE Python.
4. Gensim, sebagai LDA *library*.
5. Nltk dan Sastrawi sebagai *library corpus*, *stopwords* dan *steming* dasar Bahasa Indonesia.
6. Tweepy, sebagai API untuk mengambil data dari Twitter pada Python.

7. Matplotlib dan Bokeh, sebagai *library* untuk membuat diagram/visualisasi dari proses LDA pada {ython.
8. Openpyxl, sebagai *library* untuk menulis data ke excel

3.2.2. Bahan Penelitian

Bahan yang digunakan untuk penelitian ini adalah *tweet* dari Twitter yang membahas tentang politik Indonesia.

3.3. Diagram Alir Penelitian



Gambar 3.1 Diagram Alir Penelitian

3.3.1. Pengambilan Data

Untuk proses pengambilan data ini akan digunakan bantuan dari *library* Tweepy, yang bekerja dengan cara *men-streaming* Twitter, api ini bekerja dengan menyalin data dari Twitter. Pada proses ini peneliti membuat parameter untuk

data-data yang ingin diambil dan tidak lupa menyaring data-data yang diambil sesuai dengan apa yang dibutuhkan. Setelah data di *copy* peneliti memilih untuk meletakkan data tersebut di dalam format teks (.txt) untuk nantinya diolah lagi kedepannya. Untuk indikator-indikator yang diambil adalah:

1. *User* atau *id* dari penulis *tweet*,
2. Waktu penulisan *tweet* atau waktu *publish*,
3. *Tweet* yang berisi *maximal* 280 karakter,

akan dispesifikkan menjadi seperti berikut:

1. Mengambil data dengan bantuan *hashtag* (fitur dalam Twitter) untuk pengelompokan *tweet* dari *hashtag* #pemilu2019, #pilpres2019 dan #pileg2019 untuk jumlah masing masing data diacak.
2. Jumlah total keseluruhan data yang diambil berkisar antara 100 (seratus) hingga 6000 (enam ribu) *tweet*.
3. Proses pengambilan pada tanggal 10 Agustus 2019.

3.3.2.Preprocessing

Sebelum data diolah data terlebih dahulu menjalani beberapa tahapan agar tingkat ke akurasian data tinggi, dalam *Natural Language Processing* (NLP) data yang tidak diolah terlebih dahulu akan mengakibatkan penumpukan kosa kata yang tidak memiliki arti dianggap penting sehingga merubah hasil proses. Ditambah penggunaan Twitter yang bebas dalam menulis *tweet* didalamnya akan sangat mempengaruhi jumlah ragam kata yang dihasilkan yang berpengaruh juga pada proses LDA.

Berikut tahapan *preprocessing* yang dilakukan pada penelitian ini:

1. *Lowercase*

Pada tahapan ini data yang diperoleh (*tweet*) disamaratakan menjadi huruf kecil semua per hurufnya tanpa mengubah struktur katanya, karna jika tidak dilakukan maka contoh kata “makan” akan diartikan beda dengan “Makan”, “MAKAN” dan lain sebagainya (*case sensitive*). Pada tabel 3.1 dapat dilihat contoh penerapan sebelum dan sesudah *lowercase*.

Tabel 3.1 Contoh Penerapan Sebelum dan Sesudah *Lowercase*

Sebelum	Sesudah
Pembukaan pertama setelah renovasi lapangan sempur Ayooo semangat olahraga (at @imaginetoilet) https://path.com/p/2LKT3J	pembukaan pertama setelah renovasi lapangan sempur ayooo semangat olahraga (at @imaginetoilet) https://path.com/p/2lkt3j

2. Tokenisasi

Setelah semua huruf diubah menjadi huruf kecil semua kata-kata dibuat menjadi potongan kecil yang disebut token atau potongan kata tunggal. Tahapan ini akan menghilangkan tanda baca, angka dan karakter lainnya yang tidak ada di dalam alfabet sehingga kata akan berdiri sendiri-sendiri. Pada tabel 3.2 dapat dilihat contoh penerapan sebelum dan sesudah tokenisasi.

Tabel 3.2 Contoh Penerapan Sebelum dan Sesudah Tokenisasi

Sebelum	Sesudah
pembukaan pertama setelah renovasi lapangan sempur ayooo semangat olahraga (at @imaginetoilet) https://path.com/p/2lkt3j	‘pembukaan’ ‘pertama’ ‘setelah’ ‘renovasi’ ‘lapangan’ ‘sempur’ ‘ayooo’ ‘semangat’ ‘olahraga’ ‘at’ ‘imaginetoilet’

3. *Stopwords*

Setelah kata dipisah-pisah kata akan masuk ketahapan ini yaitu penghilangan kata yang tidak memiliki makna yang dianggap akan banyak muncul dalam hasil

sehingga akan merusak hasil. Pada tabel 3.3 dapat dilihat contoh penerapan sebelum dan sesudah *Stopwords*.

Tabel 3.3 Contoh Penerapan Sebelum dan Sesudah *Stopwords*

Sebelum	Sesudah
'pembukaan' 'pertama' 'setelah' 'renovasi' 'lapangan' 'sempur' 'ayooo' 'semangat' 'olahraga' 'at' 'imagnettoilet'	'pembukaan' 'renovasi' 'lapangan' 'sempur' 'semangat' 'olahraga'

4. *Stemming*

Pada tahapan ini kata-kata akan diolah kembali menjadi bentuk aslinya dengan cara menghilangkan imbuhan, yang bertujuan mengurangi jumlah kata yang masuk kedalam data penelitian dengan harapan meningkatkan akurasi. Pada tabel 3.4 dapat dilihat contoh penerapan sebelum dan sesudah *Stemming*.

Tabel 3.4 Contoh Penerapan Sebelum dan Sesudah *Stemming*

Sebelum	Sesudah
'pembukaan' 'renovasi' 'lapangan' 'sempur' 'semangat' 'olahraga'	'buka' 'renovasi' 'lapangan' 'sempur' 'semangat' 'olahraga'

3.3.3. Penerapan LDA

Pada pengembangannya LDA dapat digunakan untuk dua hal yaitu *generatif* dan *inferensi*. *Generatif* bertujuan untuk membuat dokumen dari kata kata yang minim (*corpus*) dan *inferensi* adalah membaca dokumen untuk dijadikan kumpulan kata-kata (*bag of word*), dan pada kesempatan kali ini akan digunakan LDA sebagai *inferensi*, dalam proses *inferensi* terbagi menjadi dua yaitu pelatihan dan proses pengujian.

1. Pelatihan menggunakan LDA Proses pelatihan menggunakan LDA bertujuan untuk mendapatkan nilai *topic proportion* dan probabilitas kata topik.

Pelatihan dilakukan dengan menggunakan aproksimasi inferensi *Gibbs Sampling*. Persamaan matematis *Gibbs Sampling* dapat dilihat pada Persamaan (1) (Kusumaningrum, 2014).

$$p(z_i = k | \vec{z}_{\sim i}, \vec{w}, a, b) = \frac{n_{k, \sim i}^{(t)} + b_t}{\sum_{t=1}^V n_{k, \sim i}^{(t)} + b_t} * n_{d, \sim i}^{(k)} + a_k \quad (1)$$

Dari Persamaan (1) dicari nilai *topic proportion* dengan menggunakan Persamaan (2) dan probabilitas kata topiknya menggunakan Persamaan (3).

$$r_{k,t} = p(w = t | z = k) = \frac{n_{t,k} + b_t}{\sum_{t=1}^V n_{t,k} + b_t} \quad (2)$$

$$q_{d,k} = p(z = k | d) = \frac{n_{d,k} + a_k}{\sum_{k=1}^V n_{d,k} + a_k} \quad (3)$$

Dengan keterangan notasi seperti pada Tabel 3.5. hasil dari pelatihan data ini akan menghasilkan kumpulan nilai untuk setiap *corpus* yang nantinya akan dijadikan panduan untuk menentukan kata apa yang akan menjadi indikasi dari dari suatu topik atau bukan.

2. Setelah didapat indikator suatu topik melalui proses pelatihan dilakukan pengujian menggunakan LDA untuk data-data sisanya proses pengujian bertujuan untuk mengukur validitas dan menghitung nilai *topic proportion* dokumen uji berdasarkan nilai probabilitas kata topik untuk setiap kata dalam dokumen uji. Pengujian dengan LDA dilakukan menggunakan Persamaan (4) (Pang dkk., 2002).

Tabel 3.5 Keterangan Notasi dan Definisi Persamaan

Notasi	Definisi
$n_{k, \sim i}^{(t)}$	Jumlah kata ke-t yang ditetapkan sebagai topik ke-k selain kata pada token ke-i
$n_{t,k}$	Jumlah kata ke-t yang ditetapkan sebagai topik ke-k pada proses <i>Gibbs Sampling</i>

$n_{d,\sim i}^{(k)}$	Jumlah kata dalam dokumen ke-d yang ditetapkan sebagai topik ke-k selain kata pada token ke-i
$n_{d,k}$	Jumlah kata dalam dokumen ke-d yang ditetapkan sebagai topik ke-k pada proses <i>Gibbs Sampling</i>
K	Jumlah topik keseluruhan
k	Topik
V	Jumlah keseluruhan kata dalam <i>corpus</i>
a_k	Hiperparameter <i>alfa</i>
b_t	Hiperparameter <i>beta</i>

$$q_{d,k} = p(z_k) * \sum_{i=1}^{N_d} r_{k,t}^{(i)} \quad (4)$$

Dimana,

$q_{d,k}$ = *topic proportion* untuk dokumen

$p(z_k)$ = probabilitas topik

$r_{k,t}^{(i)}$ = probabilitas kata topik

Setelah dilakukan pengujian menggunakan LDA dan didapatkan nilai untuk data uji, proses selanjutnya adalah melakukan perhitungan untuk mencari pola kemiripan distribusi probabilitas menggunakan *Kullback Leiber Divergence* (KLD). Rumus KLD dapat dituliskan dalam bentuk matematis seperti dalam Persamaan 5.

$$KLD = \frac{DPQ + DQP}{2} \quad (5)$$

Dimana DPQ dan DQP didapatkan menggunakan Persamaan (6) dan Persamaan (7).*

$$DPQ = \sum_{k=1}^K p_k \log\left(\frac{p_k}{q_k}\right) \quad (6)$$

$$DQP = \sum_{k=1}^K q_k \log\left(\frac{q_k}{p_k}\right) \quad (7)$$

3.3.2. Mengevaluasi Hasil LDA

Evaluasi dilakukan setelah proses pelatihan dan pengujian dengan LDA selesai dilakukan. Proses evaluasi bertujuan untuk mendapatkan hasil mengenai seberapa baik model LDA yang dibuat, dengan menghitung akurasi berdasarkan jumlah topik benar dibagi total. Perhitungan akurasi dilakukan dengan menggunakan Persamaan (8).

$$Accuracy = \frac{NB}{NB+NS} \quad (8)$$

Dimana :

NB : Total nilai benar

NS : Total nilai salah

Setelah melakukan evaluasi akan didapat persentase keberhasilan pengujian, berdasarkan rumus akurasi. peneliti menargetkan angka diatas 80% dari total jumlah pengujian dengan harapan dapat menggambarkan kinerja LDA pada topik teks pendek dalam penentuan topik politik.

BAB IV

HASIL PENELITIAN

4.1. Deskripsi Hasil Penelitian

Penelitian dilakukan dengan mengambil sumber data dari *tweet* masyarakat yang dapat diakses secara mudah melalui situs twitter.com, data tersebut akan diolah sehingga menghasilkan nilai untuk nantinya digunakan untuk menentukan topik dalam kasus yang sama, dalam hal ini teks pendek. Penelitian ini akan menghasilkan analisis dan prototipe mengenai kinerja dari algoritma LDA dalam deteksi topik pada Twitter.

4.2. Hasil Penelitian

Untuk memudahkan dalam menjelaskan hasil penelitian, berikut penjelasan hasil penelitian untuk setiap tahapannya beserta pembahasan lengkapnya.

4.2.1. Pengambilan Data

Dalam pengambilan data peneliti tidak mengambil data (*tweet*) secara manual melainkan menggunakan bantuan *library* untuk mempermudah dan mempercepat kerja, *library* yang gunakan disini adalah Tweepy. Pertama-tama sebelum menggunakan *library* dibutuhkan akses token yang berupa kode yang harus kita miliki, kode tersebut didapat dari Twitter langsung dengan cara mendaftar di halaman *developer* Twitter, developer.twitter.com, dihalaman itu *user* harus mengisi data yang diminta, proses pembuatan berlangsung 1 jam hingga 2 hari. Setelah terdaftar kita dapat meng-*generate* token yang nantinya akan digunakan untuk menggunakan tweepy. Pengambilan data dilakukan pada tanggal 10 Agustus 2019. Untuk menggunakan Tweepy kita cukup melakukan *query* dan *filtering* sesuai kebutuhan, karna pada saat penarikan data API mengirimkan

semua data (*tweet*) biasanya dalam format “json”, data tersebut tidak semua dipakai sesuai dengan yang dibutuhkan saja, berikut contoh hasil pemanggilan data *tweet* menggunakan *library* Tweepy sebelum dan sesudah *filtering* Gambar 4.1.

Sebelum filtering	sesudah filtering
<pre> created_at: 'Wed Aug 07 22:32:45 +0000 2019', 'id': 1159230927471439874, 'id_str': '1159230927471439874', 'text': '@kompascom https://t.co/MR36NeGspT', 'truncated': False, 'entities': {'hashtags': [], 'symbols': [], 'user_mentions': [{'screen_name': 'kompascom', 'name': 'Kompas.com', 'id': 23343960, 'id_str': '23343960', 'indices': [0, 10]}], 'urls': [], 'media': [{'id': 1159230918885699584, 'id_str': '1159230918885699584', 'indices': [11, 34], 'media_url': 'http://pbs.twimg.com/media/EBZqYPzU8AAYMv6.jpg', 'media_url_https': 'https://pbs.twimg.com/media/EBZqYPzU8AAYMv6.jpg', 'url': 'https://t.co/MR36NeGspT', 'display_url': 'pic.twitter.com/MR36NeGspT', 'expanded_url': 'https://twitter.com/ibnuasyary1975/status/1159230927471439 874/photo/1', 'type': 'photo', 'sizes': {'thumb': {'w': 150, 'h': 150, 'resize': 'crop'}, 'small': {'w': 680, 'h': 510, 'resize': 'fit'}, 'large': {'w': 2048, 'h': 1536, 'resize': 'fit'}, 'medium': {'w': 1200, 'h': 900, 'resize': 'fit'}}}}], 'extended_entities': {'media': </pre>	<pre> 2018-03-13 09:29:38 ibnuasyary1975 @kompascom https://t.co/MR36NeGspT </pre>

Gambar 4.1 Data Sebelum dan Sesudah *Filtering*

Dalam melakukan *filtering* dari sekian banyak data yang ada peneliti memilih hanya untuk mengambil waktu penulisan *tweet*, *user* dan *tweet*-nya hanya tiga data yaitu yang dipertahankan untuk data yang lainnya tidak digunakan. Untuk *query* yang digunakan untuk melakukan pemanggilan data adalah sebagai berikut:

1. Data diambil dengan *query*/kata kunci pilpres
2. Lokasi *tweet* dibuat adalah di Indonesia
3. *Tweet* yang diambil termasuk *retweet*

Hasil dari semua *query* tersebut menghasilkan data yang di letakkan di *file* dengan format .txt (raw.txt) dan *sheet* excel dengan nama RawData, *file* .txt

digunakan untuk proses selanjutnya dan *file* excel digunakan untuk visualisasi data agar lebih mudah dipahami dan lebih rapih.

4.2.2. PreProcessing

a) Lowercase

Proses *lowercase* (pengubahan ke bentuk non-kapital) hanya dengan menggunakan fungsi bawaan *python*, contoh kode dan contoh kasus ada pada Gambar 4.2. Untuk *file* yang digunakan adalah *raw.txt* dengan keluaran *sheet Prelowercase*.

Sebelum Lowercase	Sesudah Lowercase
RT @KPU_ID: Bergugurannya permohonan Pemohon pada sidang pembacaan putusan PHPU Pileg 2019 hari kedua Rabu (7/8) akibat tidak jelas atau ka...	rt @kpu_id: bergugurannya permohonan pemohon pada sidang pembacaan putusan phpu pileg 2019 hari kedua rabu (7/8) akibat tidak jelas atau ka...
code	
tweet=tweet.lower()	

Gambar 4.2. Proses Lowercase

b) Tokenisasi

Tokenisasi dilakukan dengan fungsi bawaan *python* dan memisahkan tiap kata menjadi berdiri sendiri, *file* yang dihasilkan dari proses ini adalah *sheet PreTokenization*. Untuk contoh dan kode pemrograman ada pada Gambar 4.3.

Sebelum Tokenisasi	Sesudah Tokenisasi
rt @kpu_id: bergugurannya permohonan pemohon pada sidang pembacaan putusan phpu pileg 2019 hari kedua rabu (7/8) akibat tidak jelas atau ka...	['rt', '@kpu_id', ':', 'bergugurannya', 'permohonan', 'pemohon', 'pada', 'sidang', 'pembacaan', 'putusan', 'phpu', 'pileg', '2019', 'hari', 'kedua', 'rabu', '(', '7/8', ')', 'akibat', 'tidak', 'jelas', 'atau', 'ka', '...']
code	
tweet =tkz.tokenize(tweet)	

Gambar 4.3. Proses Tokenisasi

c) Stopwords

Stopword dilakukan dengan bantuan *corpus* dari *library* Nltk dan tambahan *stopword* dari peneliti. Dilakukan juga penghilangan beberapa hal yang dianggap akan mengganggu nantinya seperti *hashtag*, *link* dan sebagainya, *file* yang dihasilkan dari proses ini adalah *sheet PreStopwords*. Untuk contoh dan kode pemrograman ada pada Gambar 4.4.

Sebelum Tokenisasi	Sesudah Tokenisasi
['rt', '@kpu_id', ':', 'bergugurannya', 'permohonan', 'pemohon', 'pada', 'sidang', 'pembacaan', 'putusan', 'phpu', 'pileg', '2019', 'hari', 'kedua', 'rabu', '(', '7/8', ')', 'akibat', 'tidak', 'jelas', 'atau', 'ka', '...']	['bergugurannya', 'permohonan', 'pemohon', 'sidang', 'pembacaan', 'putusan', 'phpu', 'pileg', 'rabu', 'akibat']
code	
<pre>stopwords = nltk.corpus.stopwords.words('indonesian') token = tokenfiltering(token) if re.search('[a-zA-Z]', token) and (token not in stopwords): dataoutput.append(token)</pre>	

Gambar 4.4. Proses Stopword

d) *Stemming*

Stemming dilakukan seluruhnya oleh *library* bernama Sastrawi, *file* yang dihasilkan dari proses ini adalah *sheet PreStemming*. Untuk contoh dan kode pemrograman ada pada Gambar 4.5.

Sebelum Tokenisasi	Sesudah Tokenisasi
['bergugurannya', 'permohonan', 'pemohon', 'sidang', 'pembacaan', 'putusan', 'phpu', 'pileg', 'rabu', 'akibat']	['gugur', 'mohon', 'mohon', 'sidang', 'baca', 'putus', 'phpu', 'pileg', 'rabu', 'akibat']
code	
<pre>factory = StemmerFactory() stemmer = factory.create_stemmer() hasil=stemmer.stem(data)</pre>	

Gambar 4.5. Proses Stemming

Berikut adalah waktu yang dibutuhkan untuk menyelesaikan setiap proses sesuai dengan jumlah dokumen dalam Tabel 4.1.

Tabel 4.1. Perbandingan Waktu Preprocessing per Jumlah Data

Jumlah Dokumen	Proses	Waktu yang Dibutuhkan (detik)
----------------	--------	-------------------------------

100	<i>Lowercase</i>	0.22
	Tokenisasi	0.22
	<i>Stopword</i>	0.24
	<i>Stemming</i>	24.17
1000	<i>Lowercase</i>	0.72
	Tokenisasi	0.82
	<i>Stopword</i>	1.21
	<i>Stemming</i>	138.08
6000	<i>Lowercase</i>	3.04
	Tokenisasi	3.63
	<i>Stopword</i>	5.43
	<i>Stemming</i>	454.53

4.2.3. Latent Dirichlet Allocation (LDA)

Proses pengerjaan LDA sebagian besar dikerjakan oleh Gensim *library* proses pembuatan model LDA dimulai dengan :

a) Pembuatan kamus dan *corpus*

File yang dihasilkan dari *preprocessing* (*PreStemming*) berupa *list* kata-kata bersih, kata-kata ini akan diolah sebagai bahan utama, untuk pembuatan kamus dilakukan filterasi lagi dengan ketentuan pengujian untuk 100 data menghilangkan kata yang muncul kurang dari 3 dokumen, untuk 1000 data menghilangkan kata yang muncul kurang dari 20 dokumen dan untuk 6000 data menghilangkan kata yang muncul kurang dari 120 dokumen.

Kamus merupakan kumpulan kata-kata yang dirasa unik oleh Gensim, untuk contoh dan kode pemrograman pada Gambar 4.6.

Corpus disini adalah pemberian angka random untuk setiap katanya oleh Gensim, setiap kata memiliki nomernya masing-masing, nomer ini yang nantinya akan digunakan untuk proses perhitungan lanjutan oleh LDA. Untuk contoh dan kode pemrograman pada Gambar 4.7.

Contoh Kamus
Dictionary(112 unique tokens: ['akibat', 'baca', 'gugur', 'mohon', 'phpu']...)
Kode Program
id2word = gensim.corpora.Dictionary(datafordic)
id2word.filter_extremes(no_below=2,no_above=0.9)

Gambar 4.6. Kamus yang Digunakan untuk LDA

Contoh Corpus
[[(0, 1), (1, 1), (2, 1), (3, 2), (4, 1), (5, 1), (6, 1), (7, 1), (8, 1)], [], [(9, 1), (10, 1)], [(5, 1), (11, 1), (12, 1), (13, 1), (14, 1)], [(5, 1), (11, 1), (14, 1), (15, 1), (16, 1)], [(17, 1), (18, 1), (19, 1)], [(7, 1), (20, 1), (21, 1), (22, 1), (23, 1), (24, 1), (25, 1), (26, 1), (27, 1), (28, 1)], [(0, 1), (1, 1), (2, 1), (3, 2), (4, 1), (5, 1), (6, 1), (7, 1), (8, 1)], [(0, 1), (1, 1), (2, 1), (3, 2), (4, 1), (5, 1), (6, 1), (7, 1), (8, 1)], [(0, 1), (1, 1), (2, 1), (3, 4), (4, 1), (5, 1), (6, 1), (7, 1), (8, 1), (29, 1), (30, 1)], [(16, 1), (31, 1), (32, 1)], [(18, 1), (33, 1), (34, 1), (35, 1), (36, 1), (37, 1)], [(32, 1), (38, 1), (39, 1), (40, 1), (41, 1), (42, 1), (43, 1)], [(24, 1), (38, 1), (44, 1), (45, 1), (46, 1), (47, 1)], [(5, 1), (38, 1), (48, 1), (49, 1), (50, 1), (51, 1), (52, 1)], [(38, 1), (53, 1)], [(5, 1), (11, 1), (54, 1)], [(23, 1), (24, 1), (38, 1), (55, 1), (56, 1)], [(15, 1), (19, 1), (38, 1), (43, 1), (57, 1)], [(48, 1), (51, 1), (58, 1), (59, 1), (60, 1), (61, 1), (62, 1)], [(48, 1), (51, 1), (58, 1), (59, 1), (60, 1), (61, 1), (62, 1)], [(63, 1), (64, 1), (65, 1)], [], [(19, 1), (38, 1), (48, 1), (66, 1), (67, 1), (68, 1)], [], [(18, 1), (19, 1), (69, 1), (70, 1)], [(53, 1), (71, 1)], [(6, 1), (24, 1), (51, 1), (54, 1), (65, 1), (71, 2)], [(7, 1), (20, 1), (21, 1), (22, 1), (23, 1), (24, 1), (25, 1), (26, 1), (27, 1), (28, 1)], [(9, 1), (24, 1), (38, 1), (72, 1)], [(32, 1), (39, 1), (40, 1), (41, 1), (42, 1), (43, 1), (73, 1)], [(2, 1), (11, 1), (14, 1), (24, 1), (38, 1)], [(33, 1), (34, 1), (35, 1), (36, 1), (37, 1), (60, 1), (74, 1)], [(32, 1)], [(43, 1), (60, 1), (75, 1), (76, 1), (77, 1)], [(18, 1), (26, 1), (78, 1)], [(9, 1), (18, 1), (72, 1), (79, 1)], [(24, 1), (60, 1), (70, 1), (80, 1)], [(55, 1), (81, 1), (82, 1), (83, 1)], [(16, 1), (49, 1), (84, 1)], [(53, 1)], [], [(7, 1), (20, 1), (21, 1), (22, 1), (23, 1), (24, 1), (25, 1), (26, 1), (27, 1), (28, 1)], [(45, 1), (46, 1), (85, 1), (86, 1)], [(7, 1), (20, 1), (21, 1), (22, 1), (23, 1), (24, 1), (25, 1), (26, 1), (27, 1), (28, 1)], [(19, 1), (73, 1)], [(24, 1), (38, 1), (46, 1), (86, 1)], [(15, 1), (19, 1), (57, 1), (60, 1)], [(7, 1), (20, 1), (21, 1), (22, 1), (23, 1), (24, 1), (25, 1), (26, 1), (27, 1), (28, 1)], [(19, 1), (73, 1), (87, 1)], ...]
Kode Program
corpus = [id2word.doc2bow(text) for text in corpustext]

Gambar 4.7. Corpus yang Digunakan untuk LDA

b) Pemodelan LDA

Pada tahapan ini LDA mengolah *corpus*, jumlah dokumen, jumlah topik seperti yang di gambarkan dalam visualisasi LDA Gambar 2.1., dari ketiga bahan tersebut LDA menentukan untuk representasi dari setiap topik kata-kata dan nilai dari hasil perhitungan dengan rumus pada persamaan (2) dan (3) pada bab III, pada kesempatan ini peneliti menggunakan kombinasi pengujian jumlah topik 10 dan *setting-an* lainnya *default* (*iterations*, *alpha(auto)*, *chunksize* dsb) berikut contoh hasil model LDA yang dihasilkan dan kode pemrogramannya pada Gambar 4.8. dan untuk perbandingan waktu untuk setiap percobaan terdapat pada Tabel 4.2. pemodelan juga disimpan pada sheet HasilLdaModel.

Model LDA 10 Topik
0 0.412*"menang" + 0.216*"adil" + 0.216*"curang" + 0.020*"langgar" + 0.020*"kader" + 0.020*"ulama" + 0.020*"gugat" + 0.020*"rekonsiliasi" + 0.020*"utang" + 0.020*"koruptor"
1 0.512*"adil" + 0.268*"menang" + 0.024*"langgar" + 0.024*"gugat" + 0.024*"curang" + 0.024*"kader" + 0.024*"ulama" + 0.024*"rekonsiliasi" + 0.024*"koruptor" + 0.024*"utang"
2 0.412*"menang" + 0.216*"rekonsiliasi" + 0.216*"langgar" + 0.020*"curang" + 0.020*"kader" + 0.020*"gugat" + 0.020*"adil" + 0.020*"ulama" + 0.020*"utang" + 0.020*"syariah"
3 0.580*"menang" + 0.160*"ulama" + 0.122*"kader" + 0.084*"rekonsiliasi" + 0.008*"langgar" + 0.008*"gugat" + 0.008*"curang" + 0.008*"adil" + 0.008*"utang" + 0.008*"koruptor"
4 0.524*"menang" + 0.048*"langgar" + 0.048*"gugat" + 0.048*"curang" + 0.048*"adil" + 0.048*"kader" + 0.048*"ulama" + 0.048*"utang" + 0.048*"rekonsiliasi" + 0.048*"koruptor"
5 0.307*"syariah" + 0.307*"koruptor" + 0.307*"utang" + 0.010*"menang" + 0.010*"langgar" + 0.010*"curang" + 0.010*"gugat" + 0.010*"ulama" + 0.010*"kader" + 0.010*"adil"
6 0.412*"langgar" + 0.412*"gugat" + 0.020*"menang" + 0.020*"curang" + 0.020*"kader" + 0.020*"ulama" + 0.020*"adil" + 0.020*"rekonsiliasi" + 0.020*"utang" + 0.020*"syariah"
7 0.258*"kader" + 0.258*"menang" + 0.109*"utang" + 0.109*"ulama" + 0.109*"curang" + 0.109*"langgar" + 0.010*"gugat" + 0.010*"adil" + 0.010*"rekonsiliasi" + 0.010*"koruptor"
8 0.523*"curang" + 0.048*"menang" + 0.048*"langgar" + 0.048*"gugat" + 0.048*"ulama" + 0.048*"kader" + 0.048*"adil" + 0.048*"rekonsiliasi" + 0.048*"utang" + 0.048*"koruptor"
9 0.268*"menang" + 0.268*"rekonsiliasi" + 0.268*"gugat" + 0.024*"langgar" + 0.024*"curang" + 0.024*"kader" + 0.024*"ulama" + 0.024*"adil" + 0.024*"utang" + 0.024*"koruptor"
Kode Pemrograman
lda_model = gensim.models.ldamodel.LdaModel(corpus=corpus, id2word=id2word, num_topics=int(topicnumber), alpha='auto', per_word_topics=True)

Gambar 4.8. Model LDA

Tabel 4.2. Perbandingan Waktu Pemodelan LDA

Jumlah Dokumen	Jumlah Topik	Waktu yang dibutuhkan (detik)
100	10	0.8
1000	10	1.18
6000	10	3.56

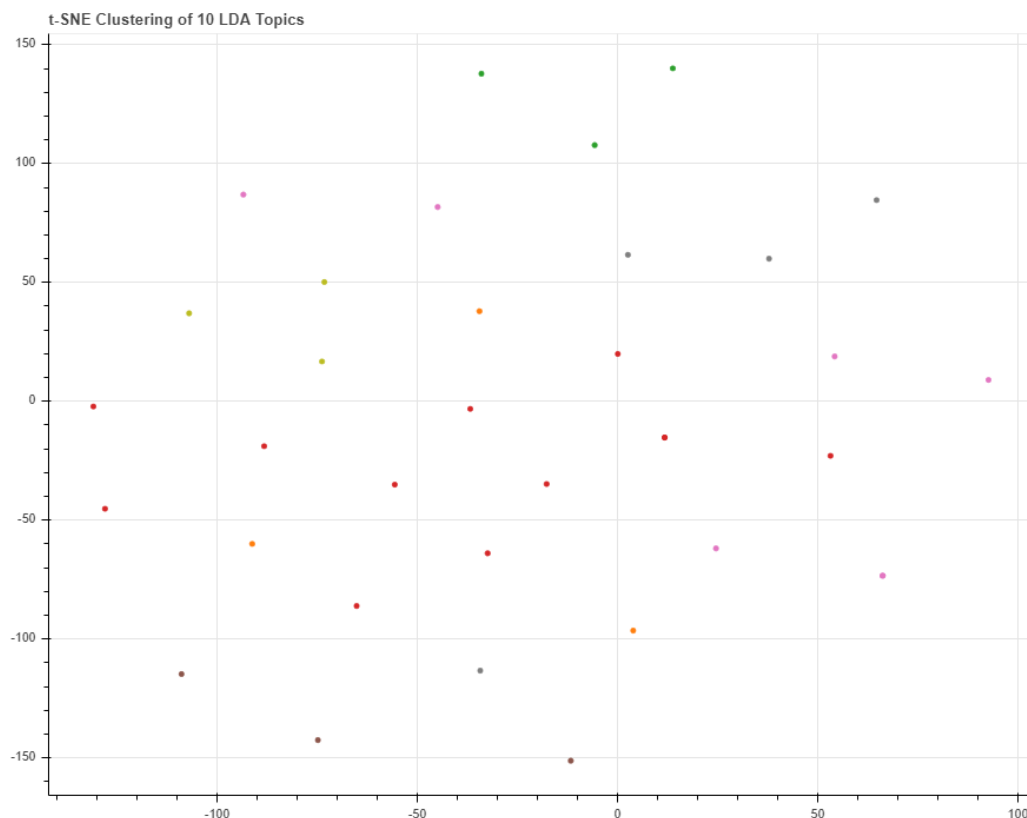
c) Penulisan hasil akhir

Setelah proses perhitungan selesai perlu dilakukan evaluasinya dan untuk mempermudah evaluasi dilakukan penulisan ulang hasil akhir agar mudah di evaluasi. Penulisan ulang ini berupa penulisan kembali data-data awal (tanggal, *user* dan *tweet* asli) ke dalam *field excel* lalu di buat dua *file* berbeda dengan nama *sheet* persentase topik dan hasil *output* dimana persentase topik berisi persentase topik mana saja dan berapa persen jumlahnya dalam satu *tweet* sedangkan hasil

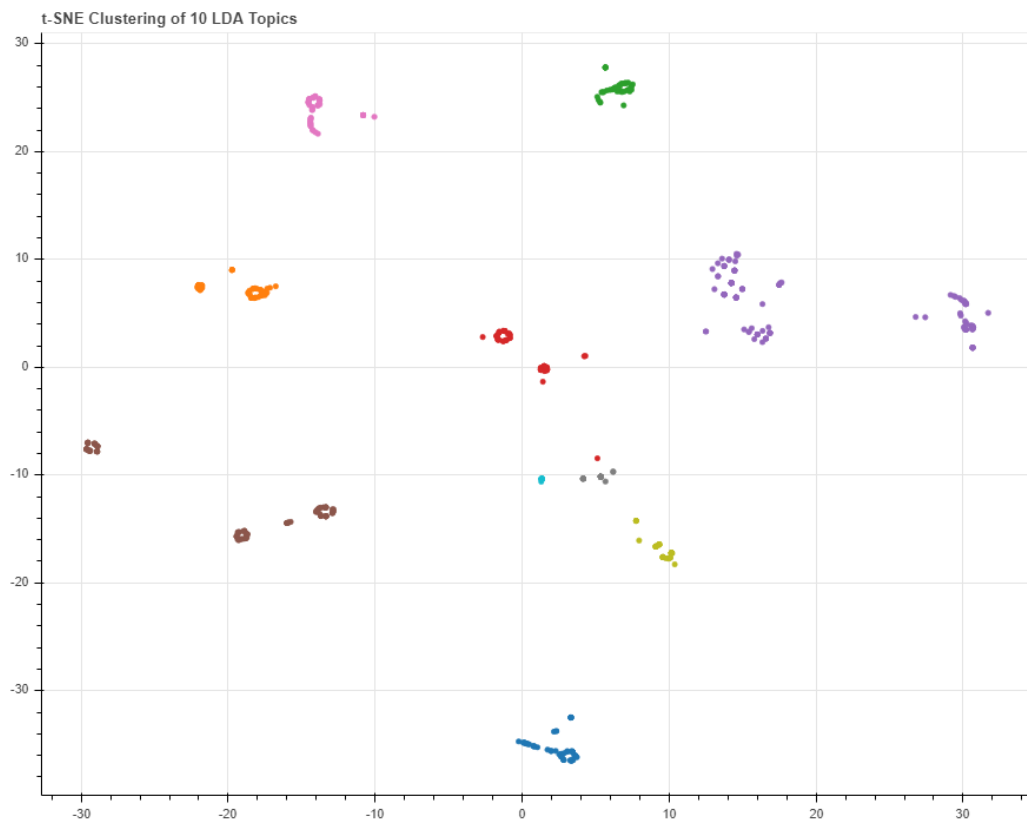
output adalah *file* yang akan di-review berisi topik utama dan *keyword* untuk setiap dokumen sehingga orang yang baca bisa lebih paham apa penentuan topik yang dihasilkan dari model LDA.

4.2.4. Visualisasi

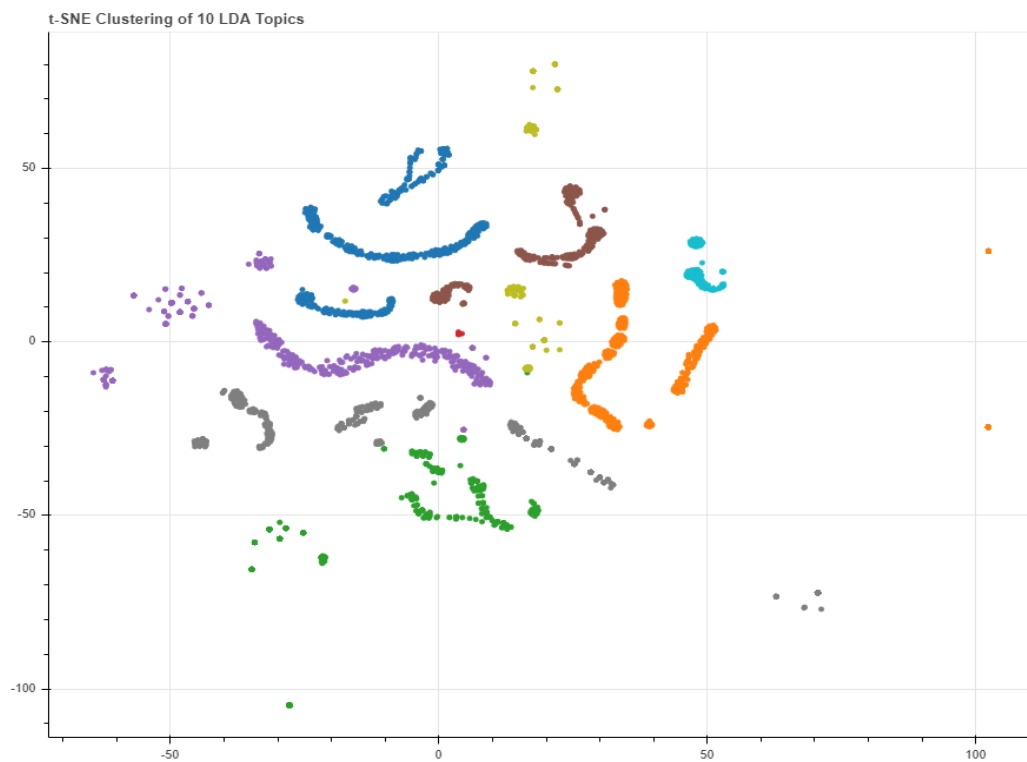
Untuk melengkapi dan mempermudah proses analisis dilakukanlah visualisasi. Proses visualisasi menggunakan bantuan *library* matplotlib proses ini berlangsung sama untuk ke tiga pengujian, dimana variabel *corpus* direpresentasikan pada sumbu X dan Y sehingga membentuk diagram dua dimensi yang menandakan posisi setiap sub kata pada topik lalu di skala agar tidak terlalu melebar, Gambar 4.9. untuk dokumen 100 Gambar 4.10. untuk dokumen 1000 dan Gambar 4.11. untuk dokumen 6000.



Gambar 4.9. Visualisasi 100 Tweet



Gambar 4.10. Visualisasi 1000 Tweet



Gambar 4.11. Visualisasi 6000 Tweet

4.2.5. Evaluasi

Tahapan ini *me-review* bagaimana kinerja penentuan topik LDA bekerja dengan menggunakan rumus akurasi sebagaimana dijelaskan pada persamaan (9) pada Bab III, berikut hasil penilaiannya Table 4.3.

Tabel 4.3. Evaluasi LDA

Jumlah Dokumen	Jumlah Topik Salah	Jumlah Topik Benar	Skor Akhir
100	10	90	0,9
1000	99	901	0,901
6000	595	5405	0,900833333

4.3. Pembahasan

Proses pendeteksian topik dimulai dengan mengambil data, proses pengambilan data menggunakan API Tweepy dengan *query* “pemilu” jumlah data yang diambil adalah total 6000 dan dari total 6000 di pisah menjadi 3 pengujian dengan pengujian-1 100 data pengujian-2 1000 data dan pengujian-3 6000 data. Untuk data yang disimpan adalah *date*, *user* dan *tweet*. Semua data disimpan di *file txt*.

File txt diolah (*preprocessing*) dengan tujuan meningkatkan keakurasian dengan membuang kata dan mengubah ke bentuk awal gambar 4.2 hingga 4.5. Tahapan-tahapan *preprocessing*nya mulai dari *lowercase* (pengembalian ke bentuk huruf nonkapital) dengan fungsi *python*, tokenisasi (pemenggalan kata) menggunakan fungsi *python*, *stopwords* (penghilangan huruf non alfabet dan beberapa kata yang dianggap akan mengganggu hasil akhir) menggunakan *stopwords* dari *library* Nltk berbahasa Indonesia dan ditambahkan *stopwords*

ciptaan peneliti, dan diakhiri dengan *stemming* (pengembalian kebentuk awal) dengan menggunakan bantuan *library* Sastrawi.

Setelah data kata-kata didapatkan proses berlanjut dengan pembuatan *corpus* oleh Gensim dengan pemberian angka per kata oleh Gensim lalu disimpan dalam list variabel *corpus*. Dari data *corpus* tersebut LDA dengan *settingan default* dan jumlah topik 10 dihasilkan model untuk 10 topik, model tersebut berisi persentase terbentuknya topik dari kata-kata pembentuknya. Lalu dilakukan penentuan topik dengan cara menghitung persentase topik untuk setiap *tweet*nya, persentase tertinggi dianggap sebagai topik utama dari suatu *tweet*, dan karena ada kemungkinan tidak ada topik yang dominan pada suatu *tweet* maka untuk meningkatkan nilai akurasi dari evaluasi maka jika ada *tweet* yang topik utamanya memiliki persentase kurang dari 0.2 dianggap tidak dapat dideteksi.

Setelah topik ditentukan secara otomatis dilakukan evaluasi, evaluasi bersifat manual dan dilakukan oleh manusia dengan cara membandingkan *tweet* asli hasil topik yang ditentukan oleh sistem, lalu dihitung jumlah benar dan salah dari semua data. Dari total keseluruhan data dari 3 kali pengujian didapatkan rata-rata kebenaran 90%.

BAB V

PENUTUP

5.1. Kesimpulan

Setelah dilakukan penelitian dengan judul Sistem Deteksi Topik Politik pada Twitter Menggunakan Algoritma *Latent Dirichlet Allocation*. Didapatkan dengan proses *preprocessing* dengan bantuan *library* Nltk untuk *stopword* dan Sastrawi untuk *stemming*, dan proses perhitungan LDA dengan bantuan Gensim dihasilkan kesimpulan bahwa dengan *settingan* dasar LDA pada *library* Gensim dan jumlah topik 10 untuk pengujian 100, 1000 dan 6000 data dihasilkan rata-rata 90% benar untuk deteksi topik LDA, nilai tersebut juga masih dapat berubah tergantung seberapa bagus *input* ataupun optimalisasi yang dilakukan pada model sehingga LDA dirasa dapat digunakan untuk mendeteksi topik pada Twitter dengan topik politik.

5.2. Saran

Penulis memiliki beberapa saran untuk peneliti serupa yang berkaitan dengan algoritma *Latent Dirichlet Allocation* untuk penentuan topik pada Twitter atau teks pendek lainnya, yaitu:

1. Untuk pengambilan data menggunakan API Tweepy lebih baik dalam waktu dimana dalam 7 hari ke belakang topik banyak dibicarakan;
2. Untuk meningkatkan akurasi *tweet* yang diambil bisa diseleksi terlebih dahulu agar sesuai yang diinginkan, bisa juga meningkatkan hasil dengan tidak mengambil *tweet* yang *me-retweet tweet* lainnya;
3. Untuk *library stopwords* dan *stemming* Bahasa Indonesia yang ada (Nltk(Indonesia)) dan Sastrawi masih sering terdapat kurang tepatnya,

alangkah baiknya menambahkan kondisi lebih/*term* pada kode pemrogramannya agar *corpus* yang dihasilkan lebih baik; dan

4. Untuk penentuan jumlah topik optimal, iterasi optimal, dan parameter lainnya dapat menggunakan *looping* dan fungsi bawaan Gensim perhitungan *coherence* dan mengambil nilai tertinggi dengan fungsi sebagai berikut, `CoherenceModel(model, text, dictionary, coherence='c_v')`.

DAFTAR PUSTAKA

- Allan, J. (2002) *Topic Detection and Tracking: Event-based Information Organization*. New York: Springer Science+Business Media.
- Blei, D.M., Y. Ng, Andrew., & I.J, Michael. (2003). "Latent Dirichlet Allocation." *Machine Learning Research Volume 3*, pp.993-1022.
- Budiardjo, M. (2007). *Dasar-Dasar Ilmu Politik*. Jakarta: PT Gramedia Pustaka Utama.
- Campbell, J.C., Hindle, A., & Stroulia, E. (2015). "Latent Dirichlet Allocation: Extracting Topics from Software Engineering Data." *The Art and Science of Analyzing Software Data*, pp.139-159.
- Fitriasih, M. & Kusumaningrum, R. (2019). "Analisis Klasifikasi Opini *Tweet* pada Media Sosial Twitter Menggunakan *Latent Dirichlet Allocation (LDA)*." *Seminar Nasional Teknologi Informasi dan Komunikasi 2019 (SENTIKA2019)*, PP.177-186
- Gil, P. (2019). "What is Twitter & How Does it Work?." *lifewire.com*. Diambil dari <https://www.lifewire.com/what-exactly-is-twitter-2483331>. Diakses pada 22 juni 2019.
- Kengken, R.I. (2014). *Pemodelan Topik untuk Media Sosial Menggunakan Latent Dirichlet Allocation Studi Kasus: Analisis Tren Berita dalam Media Sosial*. Yogyakarta: Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Gadjah Mada
- Kusumaningrum, R., Wei, H., Manurung, R. & Murni, A. (2014) "Integrated Visual Vocabulary in Latent Dirichlet Allocation-Based Scene Classification for IKONOS Image." *Journal of Applied Remote Sensing*, pp.1-18.
- Pang, B., Lee, L. & Vaithyanathan, S. (2002). "Thumbs Up ? Sentiment Classification Using Machine Learning Techniques." *Proceddings Of The Conference On Empirical Methods In Natural Language Processing (EMNLP)*, pp.79-86.
- Pratomo, Y. (2019). "Untuk Pertama Kali Twitter Ungkap Jumlah Pengguna Harian," *Kompas.com*. Diambil dari <https://tekno.kompas.com/read/2019/02/09/11340027/untuk-pertama-kali-twitter-ungkap-jumlah-pengguna-harian>. Diakses 22 April 2019.
- Rosen, A. (2017). "Tweeting Made Easier." *blog.twitter.com*. Diambil dari https://blog.twitter.com/official/en_us/topics/product/2017/tweetingmadeeasier.html. Diakses pada 22 Juni 2019.
- Saputro, D.A. (2019). *Implementasi Metode Latent Dhirichlet Allocation (LDA) untuk Menentukan Topik Teks Berita*. Kediri: Fakultas Teknik Informatika Universitas Nusantara PGRI

Lampiran-Lampiran

Diagram Penyebaran Kata Untuk 10 Topik Pada 100 *Tweet*

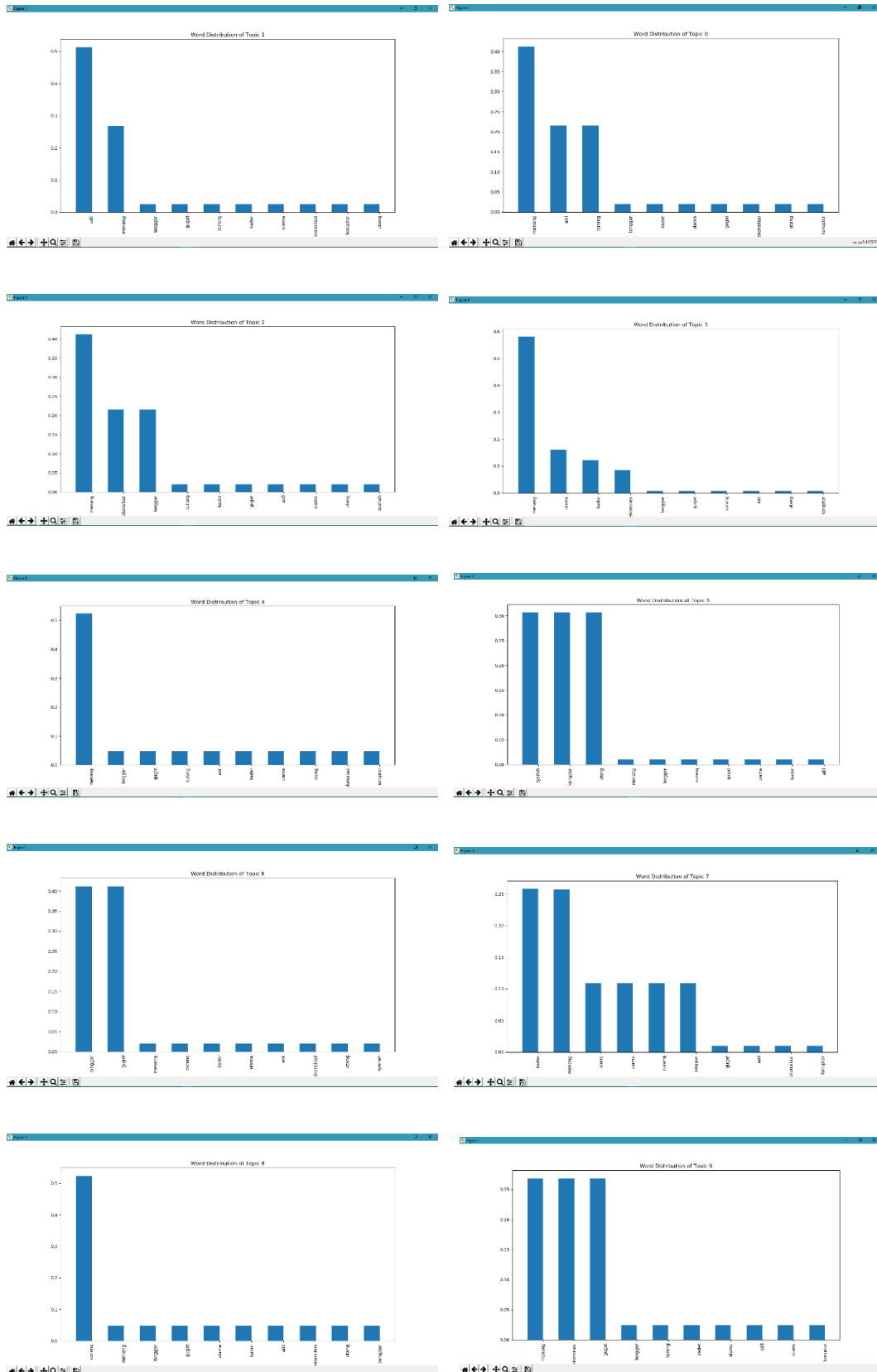


Diagram Penyebaran Kata Untuk 10 Topik Pada 1000 *Tweet*

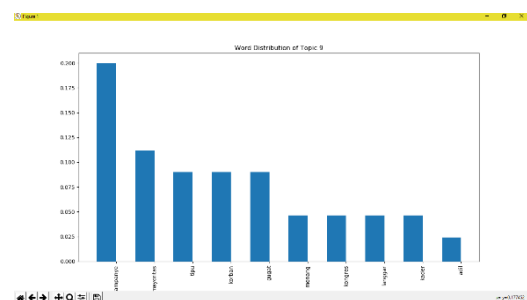
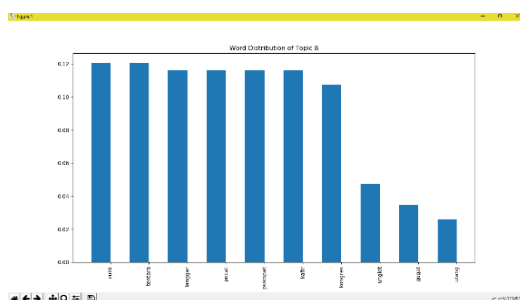
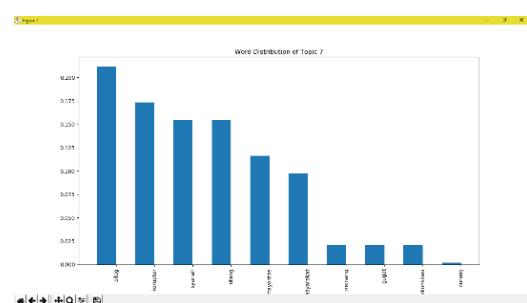
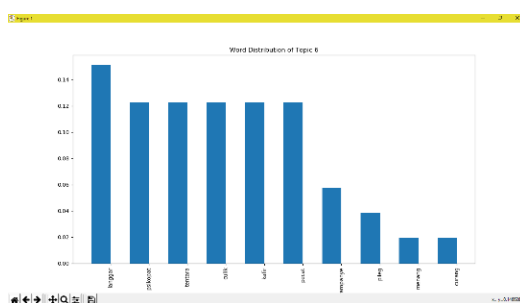
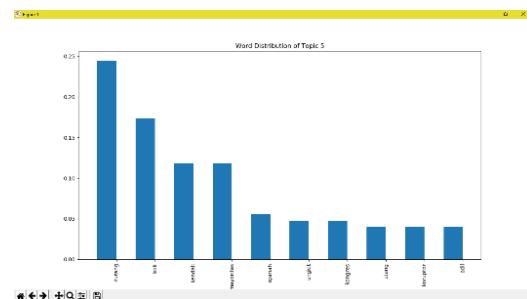
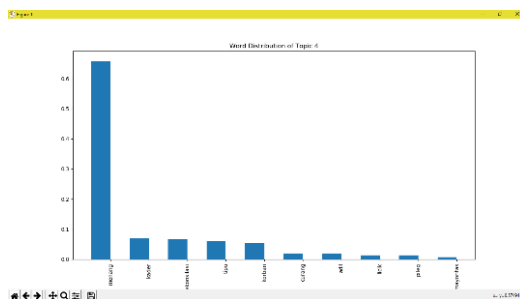
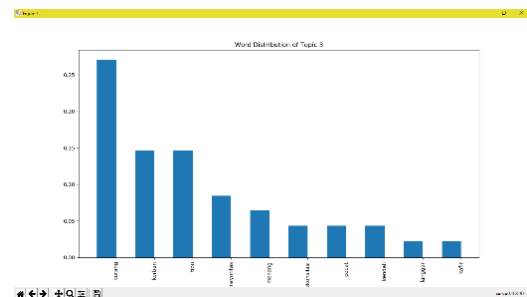
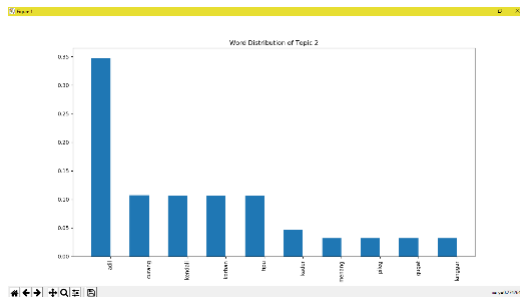
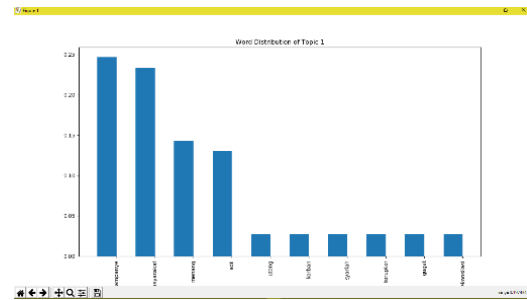
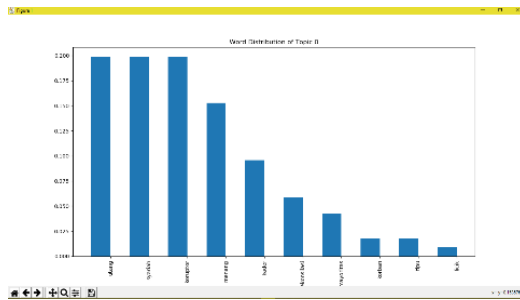
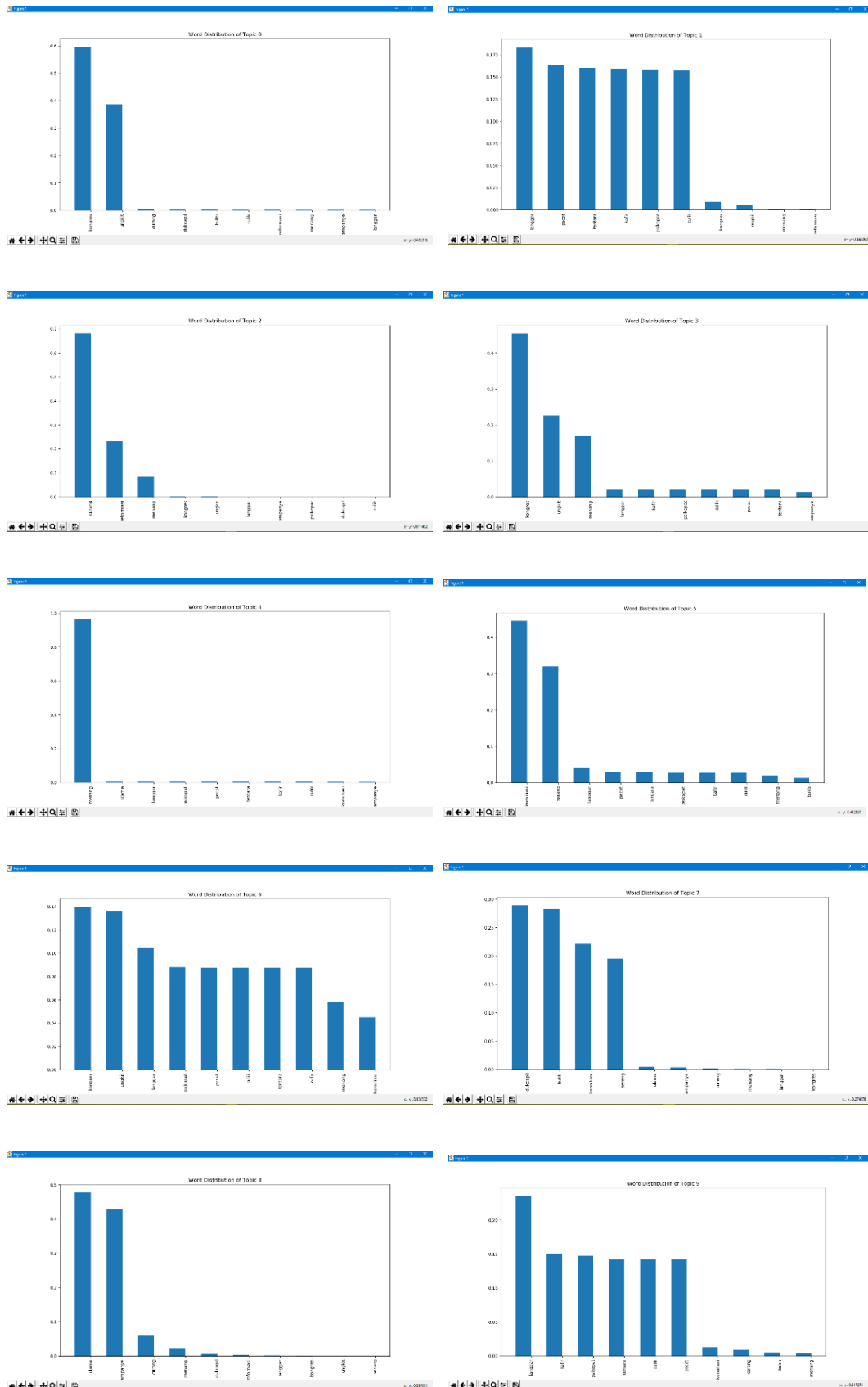


Diagram Penyebaran Kata Untuk 10 Topik Pada 6000 Tweet



Sampel 10 Tweet Dari 100 Tweet Data Asli (Rawdata) Pada Excel

Data	User	Tweet
2019-08-10 12:12:05	tinbos	RT @HukumOnline: Ragam Jenis Pelanggaran ASN Sepanjang Penyelenggaraan Pemilu 2019 https://t.co/P3NoNKISn5 https://t.co/jOuE5B3qrl
2019-08-10 12:10:18	YogaAdrianyoga6	@TeddyGusnaidi Seharusnya di pemilu kemarin itu pemerintah sadar klo ternyata yg pro Pancasila ternyata lebih besar terbukti dgn kemenangan PDIP di pileg..ini saatnya pemerintah peduli dgn suara2 anti khilafah dan menindak tegas org HTI
2019-08-10 12:08:33	HukumOnline	Ragam Jenis Pelanggaran ASN Sepanjang Penyelenggaraan Pemilu 2019 https://t.co/P3NoNKISn5 https://t.co/jOuE5B3qrl
2019-08-10 12:05:32	pejuangsosmed11	RT @Arim_Nasim: Apakah utang Negara yg ribuan trilyun Dan bunga tiap tahun diatas 200 trilyun sesuai syariah ? Apakah para Koruptor itu se...
2019-08-10 12:05:19	Wijayayanti7	RT @Arim_Nasim: Apakah utang Negara yg ribuan trilyun Dan bunga tiap tahun diatas 200 trilyun sesuai syariah ? Apakah para Koruptor itu se...
2019-08-10 12:04:37	PDIP_Wonogiri	Selamat Kepada Bapak Ir. Bambang Wuryanto, MBA. Atas Terpilihnya Menjadi Ketua DPP PDI Perjuangan Bidang Pemenangan Pemilu Tahun 2019-2024 Solid Bergerak Untuk Indonesia Raya #TetapMega #KandangBant3ng #KongresVPDIPerjuangan https://t.co/jZxkgQOKVS
2019-08-10 12:01:23	sendlaguna	Selesai sudah kontestasi pemilu, sekarang ada lagi yang dituduh "penumpang gelap".
2019-08-10 12:00:34	ibachker	@KingPurw4 @mohmahfudmd Sejak pemilu tahun kemarin Un faedah.. Semoga ke depan do beri sisa umur yang bermanfaat. Sehat lahir Dan batin
2019-08-10 11:59:01	antaranews	MK tolak gugatan caleg Partai SIRA asal Nagan Raya Aceh https://t.co/T66YsvYwdm
2019-08-10 11:58:09	lycalya	@ameliariana @riantisilvi enak bisa nyoblos bareng keluarga presiden, temen gw rumahnya di jalan veteran, pas jadi panitia pemilu bisa selfie sama bu iriana dan pak jokowi wkwkwk krn satu TPS

Sampel 10 Tweet Dari 100 Tweet Data Lowercase (Prelowercase) Pada Excel

Data	User	Tweet
2019-08-10 12:12:05	tinbos	rt @hukumonline: ragam jenis pelanggaran asn sepanjang penyelenggaraan pemilu 2019 https://t.co/p3nonklsn5 https://t.co/joue5b3qri
2019-08-10 12:10:18	YogaAdrianyoga6	@teddygusnaldi seharusnya di pemilu kemarin itu pemerintah sadar klo ternyata yg pro pancasila ternyata lebih besar terbukti dgn kemenangan pdip di pileg..ini saatnya pemerintah peduli dgn suara2 anti khilafah dan menindak tegas org hti
2019-08-10 12:08:33	HukumOnline	ragam jenis pelanggaran asn sepanjang penyelenggaraan pemilu 2019 https://t.co/p3nonklsn5 https://t.co/joue5b3qri
2019-08-10 12:05:32	pejuangsosmed11	rt @arim_nasim: apakah utang negara yg ribuan trilyun dan bunga tiap tahun diatas 200 trilyun sesuai syariah ? apakah para koruptor itu se...
2019-08-10 12:05:19	Wijayayanti7	rt @arim_nasim: apakah utang negara yg ribuan trilyun dan bunga tiap tahun diatas 200 trilyun sesuai syariah ? apakah para koruptor itu se...
2019-08-10 12:04:37	PDIP_Wonogiri	selamat kepada bapak ir. bambang wuryanto, mba. atas terpilihnya menjadi ketua dpp pdi perjuangan bidang pemenangan pemilu tahun 2019-2024 solid bergerak untuk indonesia raya #tetapmega #kandangbant3ng #kongresvpdiperjuangan https://t.co/jzxkgqokvs
2019-08-10 12:01:23	sendlaguna	selesai sudah kontestasi pemilu, sekarang ada lagi yang dituduh "penumpang gelap".
2019-08-10 12:00:34	ibachker	@kingpurw4 @mohmahfudmd sejak pemilu tahun kemarin un faedah.. semoga ke depan do beri sisa umur yang bermanfaat. sehat lahir dan batin
2019-08-10 11:59:01	antaranews	mk tolak gugatan caleg partai sira asal nagan raya aceh https://t.co/t66ysvywdm
2019-08-10 11:58:09	lycalya	@ameliariana @riantisilvi enak bisa nyoblos bareng keluarga presiden, temen gw rumahnya di jalan veteran, pas jadi panitia pemilu bisa selfie sama bu iriana dan pak jokowi wkwkwk krn satu tps

Sampel 10 Tweet Dari 100 Tweet Data Tokenisasi (*Pretokenization*) Pada Excel

Data	User	Tweet
2019-08-10 12:12:05	tinbos	['rt', '@hukumonline', ':', 'ragam', 'jenis', 'pelanggaran', 'asn', 'sepanjang', 'penyelenggaraan', 'pemilu', '2019', 'https://t.co/p3nonklsn5', 'https://t.co/joue5b3qri']
2019-08-10 12:10:18	YogaAdrianyoga6	['@teddygusnadi', 'seharusnya', 'di', 'pemilu', 'kemarin', 'itu', 'pemerintah', 'sadar', 'klo', 'ternyata', 'yg', 'pro', 'pancasila', 'ternyata', 'lebih', 'besar', 'terbukti', 'dgn', 'kemenangan', 'pdip', 'di', 'pileg', '...', 'ini', 'saatnya', 'pemerintah', 'peduli', 'dgn', 'suara', '2', 'anti', 'khilafah', 'dan', 'menindak', 'tegas', 'org', 'hti']
2019-08-10 12:08:33	HukumOnline	['ragam', 'jenis', 'pelanggaran', 'asn', 'sepanjang', 'penyelenggaraan', 'pemilu', '2019', 'https://t.co/p3nonklsn5', 'https://t.co/joue5b3qri']
2019-08-10 12:05:32	pejuangsosmed11	['rt', '@arim_nasim', ':', 'apakah', 'utang', 'negara', 'yg', 'ribuan', 'trilyun', 'dan', 'bunga', 'tiap', 'tahun', 'diatas', '200', 'trilyun', 'sesuai', 'syariah', '?', 'apakah', 'para', 'koruptor', 'itu', 'se', '...']
2019-08-10 12:05:19	Wijayayanti7	['rt', '@arim_nasim', ':', 'apakah', 'utang', 'negara', 'yg', 'ribuan', 'trilyun', 'dan', 'bunga', 'tiap', 'tahun', 'diatas', '200', 'trilyun', 'sesuai', 'syariah', '?', 'apakah', 'para', 'koruptor', 'itu', 'se', '...']
2019-08-10 12:04:37	PDIP_Wonogiri	['selamat', 'kepada', 'bapak', 'ir', ':', 'bambang', 'wuryanto', ':', 'mba', ':', 'atas', 'terpilihnya', 'menjadi', 'ketua', 'dpp', 'pdi', 'perjuangan', 'bidang', 'pemenangan', 'pemilu', 'tahun', '2019-2024', 'solid', 'bergerak', 'untuk', 'indonesia', 'raya', '#tetapmega', '#kandangbant3ng', '#kongresvpdiperjuangan', 'https://t.co/jzxkgqokvs']
2019-08-10 12:01:23	sendlaguna	['selesai', 'sudah', 'kontestasi', 'pemilu', ':', 'sekarang', 'ada', 'lagi', 'yang', 'dituduh', '...', 'penumpang', 'gelap', '...', ':']
2019-08-10 12:00:34	ibachker	['@kingpurw4', '@mohmahfudmd', 'sejak', 'pemilu', 'tahun', 'kemarin', 'un', 'faedah', '...', 'semoga', 'ke', 'depan', 'do', 'beri', 'sis', 'umur', 'yang', 'bermanfaat', ':', 'sehat', 'lahir', 'dan', 'batin']
2019-08-10 11:59:01	antaranews	['mk', 'tolak', 'gugatan', 'caleg', 'partai', 'sira', 'asal', 'nagan', 'raya', 'aceh', 'https://t.co/t66ysvywdm']
2019-08-10 11:58:09	lycalya	['@ameliariana', '@riantisilvi', 'enak', 'bisa', 'nyoblos', 'bareng', 'keluarga', 'presiden', ':', 'temen', 'gw', 'rumahnya', 'di', 'jalan', 'veteran', ':', 'pas', 'jadi', 'panitia', 'pemilu', 'bisa', 'selfie', 'sama', 'bu', 'iriana', 'dan', 'pak', 'jokowi', 'wkwwk', 'krn', 'satu', 'tps']

Sampel 10 Tweet Dari 100 Tweet Data Stopwords (Prestopwords) Pada Excel

Data	User	Tweet
2019-08-10 12:12:05	tinbos	['ragam', 'jenis', 'pelanggaran', 'penyelenggaraan']
2019-08-10 12:10:18	YogaAdrianyoga6	['pemerintah', 'pancasila', 'terbukti', 'kemenangan', 'pileg', 'pemerintah', 'peduli', 'anti', 'khilafah', 'menindak']
2019-08-10 12:08:33	HukumOnline	['ragam', 'jenis', 'pelanggaran', 'penyelenggaraan']
2019-08-10 12:05:32	pejuangsosmed11	['utang', 'ribuan', 'diatas', 'syariah', 'koruptor']
2019-08-10 12:05:19	Wijayayanti7	['utang', 'ribuan', 'diatas', 'syariah', 'koruptor']
2019-08-10 12:04:37	PDIP_Wonogiri	['terpilihnya', 'perjuangan', 'bidang', 'pemenangan', 'solid', 'bergerak', 'raya']
2019-08-10 12:01:23	sendlaguna	['kontestasi', 'dituduh', 'penumpang', 'gelap']
2019-08-10 12:00:34	ibachker	['faedah', 'semoga', 'sisa', 'umur', 'bermanfaat', 'sehat', 'lahir', 'batin']
2019-08-10 11:59:01	antaranews	['tolak', 'gugatan', 'sira', 'nagan', 'raya']
2019-08-10 11:58:09	lycalya	['nyoblos', 'bareng', 'keluarga', 'temen', 'rumahnya', 'veteran', 'panitia', 'selfie', 'wkwwk']

Sampel 10 Tweet Dari 100 Tweet Data Stemming (Prestemming) Pada Excel

Data	User	Tweet
2019-08-10 12:12:05	tinbos	['ragam', 'jenis', 'langgar']
2019-08-10 12:10:18	YogaAdrianyoga6	['pancasila', 'bukti', 'menang', 'pileg', 'peduli', 'anti', 'khilafah', 'tindak']
2019-08-10 12:08:33	HukumOnline	['ragam', 'jenis', 'langgar']
2019-08-10 12:05:32	pejuangsosmed11	['utang', 'syariah', 'koruptor']
2019-08-10 12:05:19	Wijayayanti7	['utang', 'syariah', 'koruptor']
2019-08-10 12:04:37	PDIP_Wonogiri	['bidang', 'menang', 'solid', 'gerak', 'raya']
2019-08-10 12:01:23	sendlaguna	['kontestasi', 'tuduh', 'tumpang', 'gelap']
2019-08-10 12:00:34	ibachker	['faedah', 'moga', 'sisa', 'umur', 'manfaat', 'sehat', 'lahir', 'batin']
2019-08-10 11:59:01	antaranews	['tolak', 'gugat', 'sira', 'nagan', 'raya']
2019-08-10 11:58:09	lycalya	['nyoblos', 'bareng', 'keluarga', 'temen', 'rumah', 'veteran', 'panitia', 'selfie', 'wkwwk']

Sampel Corpus Untuk 100 Tweet

[[(0, 1), [(1, 1)], [(0, 1)], [(2, 1), (3, 1), (4, 1)], [(2, 1), (3, 1), (4, 1)], [(1, 1)], [], [(5, 1)], [], [], [(2, 1), (3, 1), (4, 1)], [], [], [(1, 1), (6, 1)], [(1, 1), (7, 1)], [(8, 1)], [], [(1, 1)], [], [], [(1, 1), (6, 1)], [], [], [(8, 1)], [], [], [(1, 2)], [], [(1, 1), (7, 1)], [(1, 1)], [], [(0, 1)], [], [], [(9, 1)], [(10, 1)], [], [(0, 1)], [], [(10, 1)], [(1, 1)], [(4, 1), (9, 1)], [], [(1, 1), (10, 1)], [(1, 1), (7, 1)], [], [(9, 1)], [], [(1, 1)], [(5, 1)], [(5, 1)], [], [(8, 1)], [(1, 1), (6, 1)], [], [(7, 1)], [], [(1, 1)], [(1, 1)], [], []]]

Sampel LDA Model Untuk 100 Tweet Dengan 10 Topik

0 0.412*"menang" + 0.216*"adil" + 0.216*"curang" + 0.020*"langgar" + 0.020*"kader" + 0.020*"ulama" + 0.020*"gugat" + 0.020*"rekonsiliasi" + 0.020*"utang" + 0.020*"koruptor"
1 0.512*"adil" + 0.268*"menang" + 0.024*"langgar" + 0.024*"gugat" + 0.024*"curang" + 0.024*"kader" + 0.024*"ulama" + 0.024*"rekonsiliasi" + 0.024*"koruptor" + 0.024*"utang"
2 0.412*"menang" + 0.216*"rekonsiliasi" + 0.216*"langgar" + 0.020*"curang" + 0.020*"kader" + 0.020*"gugat" + 0.020*"adil" + 0.020*"ulama" + 0.020*"utang" + 0.020*"syariah"
3 0.580*"menang" + 0.160*"ulama" + 0.122*"kader" + 0.084*"rekonsiliasi" + 0.008*"langgar" + 0.008*"gugat" + 0.008*"curang" + 0.008*"adil" + 0.008*"utang" + 0.008*"koruptor"
4 0.524*"menang" + 0.048*"langgar" + 0.048*"gugat" + 0.048*"curang" + 0.048*"adil" + 0.048*"kader" + 0.048*"ulama" + 0.048*"utang" + 0.048*"rekonsiliasi" + 0.048*"koruptor"
5 0.307*"syariah" + 0.307*"koruptor" + 0.307*"utang" + 0.010*"menang" + 0.010*"langgar" + 0.010*"curang" + 0.010*"gugat" + 0.010*"ulama" + 0.010*"kader" + 0.010*"adil"
6 0.412*"langgar" + 0.412*"gugat" + 0.020*"menang" + 0.020*"curang" + 0.020*"kader" + 0.020*"ulama" + 0.020*"adil" + 0.020*"rekonsiliasi" + 0.020*"utang" + 0.020*"syariah"
7 0.258*"kader" + 0.258*"menang" + 0.109*"utang" + 0.109*"ulama" + 0.109*"curang" + 0.109*"langgar" + 0.010*"gugat" + 0.010*"adil" + 0.010*"rekonsiliasi" + 0.010*"koruptor"
8 0.523*"curang" + 0.048*"menang" + 0.048*"langgar" + 0.048*"gugat" + 0.048*"ulama" + 0.048*"kader" + 0.048*"adil" + 0.048*"rekonsiliasi" + 0.048*"utang" + 0.048*"koruptor"
9 0.268*"menang" + 0.268*"rekonsiliasi" + 0.268*"gugat" + 0.024*"langgar" + 0.024*"curang" + 0.024*"kader" + 0.024*"ulama" + 0.024*"adil" + 0.024*"utang" + 0.024*"koruptor"

Sampel 10 Tweet Dari 100 Tweet Persentase Topik (PersentaseTopik) Pada Excel

Date	User	Tweet	Persentase
2019-08-10 12:12:05	tinbos	RT @HukumOnline: Ragam Jenis Pelanggaran ASN Sepanjang Penyelenggaraan Pemilu 2019 https://t.co/P3NoNKISn5 https://t.co/jOuE5B3qrl	[(0, 0.049552433), (1, 0.049507465), (2, 0.04956154), (3, 0.052606147), (4, 0.048519056), (5, 0.049711965), (6, 0.5513841), (7, 0.05158049), (8, 0.048519053), (9, 0.049057744)]
2019-08-10 12:10:18	YogaAdrianyoga6	@TeddyGusnadi Seharusnya di pemilu kemarin itu pemerintah sadar klo ternyata yg pro Pancasila ternyata lebih besar terbukti dgn kemenangan PDIP di pileg..ini saatnya pemerintah peduli dgn suara2 anti khilafah dan menindak tegas org HTI	[(0, 0.049565148), (1, 0.049513865), (2, 0.049565077), (3, 0.55394095), (4, 0.04853314), (5, 0.049711954), (6, 0.050000787), (7, 0.051586434), (8, 0.04851904), (9, 0.04906357)]
2019-08-10 12:08:33	HukumOnline	Ragam Jenis Pelanggaran ASN Sepanjang Penyelenggaraan Pemilu 2019 https://t.co/P3NoNKISn5 https://t.co/jOuE5B3qrl	[(0, 0.049552433), (1, 0.049507465), (2, 0.049561463), (3, 0.052606147), (4, 0.048519056), (5, 0.049711965), (6, 0.5513842), (7, 0.05158049), (8, 0.048519053), (9, 0.049057744)]
2019-08-10 12:05:32	pejuangsosmed11	RT @Arim_Nasim: Apakah utang Negara yg ribuan trilyun Dan bunga tiap tahun diatas 200 trilyun sesuai syariah ? Apakah para Koruptor itu se...	[(0, 0.024741618), (1, 0.024719168), (2, 0.024741631), (3, 0.026266344), (4, 0.024225652), (5, 0.7758676), (6, 0.02496549), (7, 0.025752254), (8, 0.02422565), (9, 0.02449462)]
2019-08-10 12:05:19	Wijayayanti7	RT @Arim_Nasim: Apakah utang Negara yg ribuan trilyun Dan bunga tiap tahun diatas 200 trilyun sesuai syariah ? Apakah para Koruptor itu se...	[(0, 0.024741616), (1, 0.024719164), (2, 0.02474163), (3, 0.026266342), (4, 0.024225648), (5, 0.7758674), (6, 0.024965487), (7, 0.02575246), (8, 0.024225647), (9, 0.024494618)]
2019-08-10 12:04:37	PDIP_Wonogiri	Selamat Kepada Bapak Ir. Bambang Wuryanto, MBA. Atas Terpilihnya Menjadi Ketua DPP PDI Perjuangan Bidang Pemenangan Pemilu Tahun 2019-2024 Solid Bergerak Untuk Indonesia Raya #TetapMega #KandangBant3ng #KongresVPDI Perjuangan https://t.co/jZxkgQOKVS	[(0, 0.049565464), (1, 0.049513854), (2, 0.049565032), (3, 0.5539433), (4, 0.04853061), (5, 0.049711958), (6, 0.05000079), (7, 0.051586404), (8, 0.048519045), (9, 0.049063556)]
2019-08-10 12:01:23	sendlaguna	Selesai sudah kontestasi pemilu, sekarang ada lagi yang dituduh "penumpang gelap".	[(0, 0.09938279), (1, 0.0992926), (2, 0.09938284), (3, 0.105507344), (4, 0.09731024), (5, 0.099702746), (6, 0.100282036), (7, 0.10343854), (8, 0.09731023), (9, 0.09839064)]
2019-08-10 12:00:34	ibachker	@KingPurw4 @mohmahfudmd Sejak pemilu tahun kemarin Un faedah.. Semoga ke depan do beri sisa umur yang bermanfaat. Sehat lahir Dan batin	[(0, 0.09938279), (1, 0.0992926), (2, 0.09938284), (3, 0.105507344), (4, 0.09731024), (5, 0.099702746), (6, 0.100282036), (7, 0.10343854), (8, 0.09731023), (9, 0.09839064)]
2019-08-10 11:59:01	antaranews	MK tolak gugatan caleg Partai SIRA asal Nagan Raya Aceh https://t.co/T66YsvYwdm	[(0, 0.049552433), (1, 0.049507465), (2, 0.04955246), (3, 0.052606147), (4, 0.048519056), (5, 0.049711965), (6, 0.5513896), (7, 0.05157464), (8, 0.048519053), (9, 0.049067184)]
2019-08-10 11:58:09	lycalya	@ameliariana @riantisilvi enak bisa nyoblos bareng keluarga presiden, temen gw rumahnya di jalan veteran, pas jadi panitia pemilu bisa selfie sama bu iriana dan pak jokowi wkwkwk krn satu TPS	[(0, 0.09938279), (1, 0.0992926), (2, 0.09938284), (3, 0.105507344), (4, 0.09731024), (5, 0.099702746), (6, 0.100282036), (7, 0.10343854), (8, 0.09731023), (9, 0.09839064)]

Sampel 10 Tweet Dari 100 Tweet Evaluasi (100) Pada Excel Hasil Evaluasi

Tweet	Main topik	Keyword	T/F
RT @HukumOnline: Ragam Jenis Pelanggaran ASN Sepanjang Penyelenggaraan Pemilu 2019 https://t.co/P3NoNKISn5 https://t.co/jOuE5B3qrl	6	langgar, gugat, menang, curang, kader, ulama, adil, rekonsiliasi, utang, syariah	T
@TeddyGusnaldi Seharusnya di pemilu kemarin itu pemerintah sadar klo ternyata yg pro Pancasila ternyata lebih besar terbukti dgn kemenangan PDIP di pileg..ini saatnya pemerintah peduli dgn suara2 anti khilafah dan menindak tegas org HTI	3	menang, ulama, kader, rekonsiliasi, langgar, gugat, curang, adil, utang, koruptor	T
Ragam Jenis Pelanggaran ASN Sepanjang Penyelenggaraan Pemilu 2019 https://t.co/P3NoNKISn5 https://t.co/jOuE5B3qrl	6	langgar, gugat, menang, curang, kader, ulama, adil, rekonsiliasi, utang, syariah	T
RT @Arim_Nasim: Apakah utang Negara yg ribuan trilyun Dan bunga tiap tahun diatas 200 trilyun sesuai syariah ? Apakah para Koruptor itu se...	5	syariah, koruptor, utang, menang, langgar, curang, gugat, ulama, kader, adil	T
RT @Arim_Nasim: Apakah utang Negara yg ribuan trilyun Dan bunga tiap tahun diatas 200 trilyun sesuai syariah ? Apakah para Koruptor itu se...	5	syariah, koruptor, utang, menang, langgar, curang, gugat, ulama, kader, adil	T
Selamat Kepada Bapak Ir. Bambang Wuryanto, MBA. Atas Terpilihnya Menjadi Ketua DPP PDI Perjuangan Bidang Pemenangan Pemilu Tahun 2019-2024 Solid Bergerak Untuk Indonesia Raya #TetapMega #KandangBant3ng #KongresVPDI Perjuangan https://t.co/jZxkgQOKVS	3	menang, ulama, kader, rekonsiliasi, langgar, gugat, curang, adil, utang, koruptor	F
Selesai sudah kontestasi pemilu, sekarang ada lagi yang dituduh "penumpang gelap".	-1	['null']	T
@KingPurw4 @mohmahfudmd Sejak pemilu tahun kemarin Un faedah.. Semoga ke depan do beri sisa umur yang bermanfaat. Sehat lahir Dan batin	-1	['null']	T
MK tolak gugatan caleg Partai SIRA asal Nagan Raya Aceh https://t.co/T66YsvYwdm	6	langgar, gugat, menang, curang, kader, ulama, adil, rekonsiliasi, utang, syariah	T
@ameliariana @riantisilvi enak bisa nyoblos bareng keluarga presiden, temen gw rumahnya di jalan veteran, pas jadi panitia pemilu bisa selfie sama bu iriana dan pak jokowi wkwkwk krn satu TPS	-1	['null']	T

DAFTAR RIWAYAT HIDUP



Penulis bernama lengkap Khairul Hudha Nasution. Dilahirkan di Bekasi pada tanggal 14 Desember 1997 dari pasangan Kholad Nasution dan Kholilah Pulungan. Penulis adalah anak pertama dari tiga bersaudara. Penulis berkebangsaan Indonesia dan selama kuliah tinggal bersama orangtua.

Penulis menempuh pendidikan formal di SDN Sepanjang Jaya VIII (2003 – 2009), SMP IT Al-Hassan (2009 – 2012), SMA IT AS-Syifa (2012 – 2014) dan SMAN 1 Tambun Selatan (2014 – 2015). Pada Tahun 2015, penulis lulus seleksi masuk Universitas Negeri Jakarta melalui jalur PENMABA dan diterima di Fakultas Teknik, Program Studi Pendidikan Teknik Informatika dan Komputer dengan mengambil konsentrasi peminatan Rekayasa Perangkat Lunak (RPL) dan lulus pada tahun 2019.

Penulis pernah mengikuti program Praktik Kerja Lapangan (PKL) di PT. Bentang Mitrag Guna pada tahun 2018 dan mengikuti program Praktik Keterampilan Mengajar (PKM) di SMK N 26 Jakarta Timur pada semester ganjil tahun ajaran 2018/2019. Dalam menyelesaikan studi, penulis mengadakan penelitian tugas akhir skripsi dengan judul “Sistem Deteksi Topik Politik pada Twitter Menggunakan Algoritma *Latent Dirichlet Allocation*.” untuk mendapat gelar sarjana.