

***TOPIC MODELING KASUS PENIPUAN DI TWITTER  
MENGUNAKAN METODE *LATENT DIRICHLET  
ALLOCATION (LDA)****



Disusun Oleh:

N a m a : Sallu Muharomah

NIM : 19523223

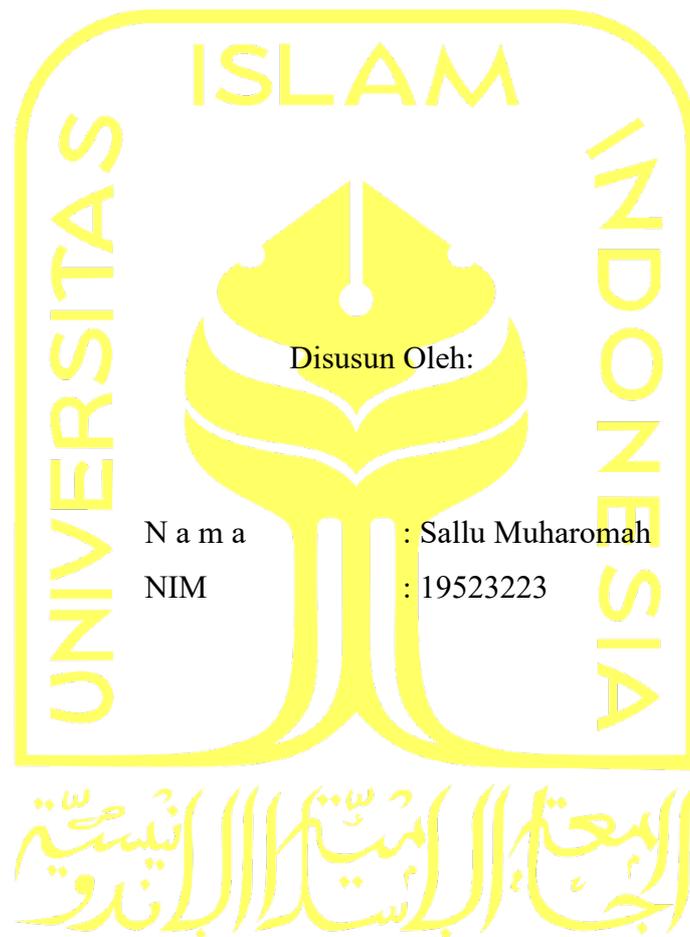
**PROGRAM STUDI INFORMATIKA – PROGRAM SARJANA  
FAKULTAS TEKNOLOGI INDUSTRI  
UNIVERSITAS ISLAM INDONESIA**

**2023**

HALAMAN PENGESAHAN DOSEN PEMBIMBING

***TOPIC MODELING KASUS PENIPUAN DI TWITTER  
MENGUNAKAN METODE LATENT DIRICHLET  
ALLOCATION (LDA)***

**TUGAS AKHIR**



Yogyakarta, 9 Juli 2023

Pembimbing,

( Chanifah Indah Ratnasari, S.Kom., M.Kom. )

HALAMAN PENGESAHAN DOSEN PENGUJI

**TOPIC MODELING KASUS PENIPUAN DI TWITTER  
MENGUNAKAN METODE *LATENT DIRICHLET  
ALLOCATION (LDA)***

**TUGAS AKHIR**

Telah dipertahankan di depan sidang pengujian sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer dari Program Studi Informatika – Program Sarjana di Fakultas Teknologi Industri Universitas Islam Indonesia

Yogyakarta, 9 Juli 2023

Tim Penguji

Chanifah Indah Ratnasari, S.KOM., M.KOM.

**Anggota 1**

Arrie Kurniawardhani, S.SI., M.KOM.

**Anggota 2**

Aridhanyati Arifin, S.T., M.CS.

Mengetahui,

Ketua Program Studi Informatika – Program Sarjana

Fakultas Teknologi Industri

Universitas Islam Indonesia



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

**HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR**

iv

**HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR**

Yang bertanda tangan di bawah ini:

Nama : Sallu Muharomah

NIM : 19523223

Tugas akhir dengan judul:

***TOPIC MODELING KASUS PENIPUAN DI TWITTER  
MENGUNAKAN METODE *LATENT DIRICHLET  
ALLOCATION (LDA)****

Menyatakan bahwa seluruh komponen dan isi dalam tugas akhir ini adalah hasil karya saya sendiri. Apabila di kemudian hari terbukti ada beberapa bagian dari karya ini adalah bukan hasil karya sendiri, tugas akhir yang diajukan sebagai hasil karya sendiri ini siap ditarik kembali dan siap menanggung risiko dan konsekuensi apapun.

Demikian surat pernyataan ini dibuat, semoga dapat dipergunakan sebagaimana mestinya.

Yogyakarta, 1 Juli 2023



(Sallu Muharomah)

## **HALAMAN PERSEMBAHAN**

Skripsi ini saya persembahkan untuk keempat orang tua saya. Berkat dukungan, ridho, perhatian, serta doa dari beliau saya dapat menyelesaikan tanggung jawab saya sehingga mendapatkan gelar S.Kom ini. Dan saya persembahkan untuk adik saya tersayang agar memotivasi dia untuk dapat mengejar cita-cita setinggi mungkin.

**HALAMAN MOTO**

*The scary news is you're on your own now, but the cool news is you're on your own now.*

(Taylor Alison Swift)

## KATA PENGANTAR

*Assalamu'alaikum Wr. Wb.*

Dengan rendah hati dan penuh rasa syukur kehadirat Allah Swt atas segala limpahan kasih, karunia, dan kehendak-Nya sehingga Tugas Akhir Skripsi dengan Judul *Topic Modeling* Kasus Penipuan di Twitter Menggunakan Metode *Latent Dirichlet Allocation* (LDA), dapat diselesaikan dengan baik. Selesaiannya Tugas Akhir ini tidak terlepas dari bantuan, bimbingan dan doa dari berbagai pihak. Pada kesempatan ini ingin disampaikan terima kasih kepada semua pihak yang telah membantu dalam pembuatan karya ini, ucapan terima kasih yang sebesar-besarnya kepada yang terhormat:

1. Bapak Prof. Fathul Wahid, S.T.,M.Sc.,Ph.D selaku Rektor Universitas Islam Indonesia.
2. Bapak Hari Purnomo, Prof., Dr., Ir., M.T., IPU, ASEAN.Eng selaku Dekan Fakultas Teknologi Industri.
3. Bapak Dhomas Hatta Fudholi, S.T., M.Eng., Ph.D. selaku Kepala Prodi Informatika.
4. Ibu Chanifah Indah Ratnasari, S.Kom., M.Kom. selaku dosen pembimbing yang telah meluangkan waktu selama proses bimbingan.
5. Papah Cimpi Mbongo dan Mamah Serly Ramadani selaku orang tua yang selalu mendukung secara materi maupun non-materi, memberikan doa dan restu sehingga Tugas Akhir ini dapat diselesaikan dengan baik.
6. Ahmad Zulfikar Assalam selaku adik yang selalu menghibur dan memberi semangat sehingga Tugas Akhir ini dapat dikerjakan dengan menyenangkan.
7. Daddy Abdellah dan Bunda Fitri selaku orang tua yang selalu mendukung secara materi maupun non-materi, memberikan doa dan restu sehingga Tugas Akhir ini dapat diselesaikan dengan baik.
8. Any, Pakde, Alm.Ambo Rizal, Ambo Hadi, Ibu Agus, Om Dede, Tante Wina, Ka Sisi, Ka Taris, Ka Cindy, Ka Terry, Ka Cania, Sasa, Cilak, dan Acabi selaku keluarga yang selalu memberikan dukungan dan semangat sehingga saya dapat menyelesaikan Tugas Akhir ini dengan baik.
9. Fariz, Mba Ratu, Ciko, Adis, Rania, Rima, Fardhan, Rapi, Reizi, Rehan, Athiya, Rama, Mutmaw, Dhy, Mba Nad, dan Tijes selaku teman-teman saya yang selalu

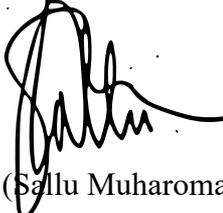
membuat hari saya berwarna sehingga saya bersemangat untuk mengerjakan Tugas Akhir ini.

10. Teman-teman seperjuangan Informatika Angkatan 2019 yang senantiasa berjuang untuk memperoleh gelar S.Kom, yang telah memberikan pengalaman yang tidak akan terlupakan selama kuliah.
11. Terakhir, untuk diri saya sendiri Sallu Muharomah yang telah berjuang dan selalu bersemangat untuk memperjuangkan gelar S.Kom ini.

Penulis menyadari bahwa dalam penulisan Tugas Akhir ini masih belum sempurna. Harapan penulis semoga tugas akhir ini dapat bermanfaat untuk penulis sendiri dan juga pembaca. Akhir kata, semoga Allah SWT. Selalu melimpahkan nikmat dan rahmat-Nya kepada kita semua. Aamiin aamiin ya robbal'alamin.

*Wassalamualaikum Wr. Wb.*

Yogyakarta, 1 Juli 2023



(Sallu Muharomah)

## SARI

Penipuan merupakan fenomena yang terus eksis dalam masyarakat dengan modus operandi yang terus berkembang seiring perkembangan zaman. Modus operandi penipuan senantiasa berubah dan berkembang seiring dengan kemajuan teknologi, globalisasi, dan pergeseran perilaku konsumen. Dalam era digital saat ini, media sosial memiliki peran signifikan dalam penyebaran informasi mengenai penipuan. Twitter, sebagai salah satu platform media sosial yang banyak digunakan, memberikan akses mudah dan cepat terhadap informasi yang relevan. Oleh karena itu, penelitian tentang modus operandi penipuan yang tersebar di media sosial, terutama di Twitter, penting untuk dilakukan. Penelitian ini melakukan pemodelan topik dengan menggunakan metode *Latent Dirichlet Allocation* (LDA) untuk mengklasifikasikan dan mengidentifikasi topik-topik penipuan yang sering dibahas oleh pengguna Twitter di Indonesia. Dengan menerapkan pemodelan LDA, penelitian ini bertujuan untuk memahami lebih komprehensif mengenai topik-topik penipuan yang sering muncul di Twitter. Berdasarkan penelitian yang dilakukan, terdapat dua penelitian menggunakan *dataset* yang berbeda, sehingga didapatkan 10 model dengan nilai koheren sebesar 0.600753 pada iterasi pertama dan pada iterasi kedua dihasilkan 9 model dengan nilai koheren sebesar 0.587142.

Kata kunci: Penipuan; Pemodelan Topik; Twitter; LDA

## GLOSARIUM

- Coherence Score* Digunakan untuk mengukur nilai suatu topik dengan melalui pengukuran tingkat kesamaan semantik dalam kata-kata yang terdapat pada topik.
- Preprocessing* Tahapan untuk menghilangkan beberapa permasalahan yang dapat mengganggu saat pemrosesan data.
- Data scraping* Teknik atau metode otomatisasi yang memungkinkan seseorang untuk mengekstrak data dari sebuah *website*, *database*, aplikasi *enterprise*, atau sistem *legacy* yang kemudian dapat menyimpannya ke dalam sebuah file dengan format tabular atau *spreadsheet*.
- Topic modeling* Suatu pendekatan untuk menganalisis kumpulan dokumen berbentuk teks dan mengelompokkan menjadi beberapa topik.
- Gensim Sebuah *library* dalam bahasa pemrograman Python yang digunakan untuk memodelkan dan memproses teks, terutama untuk tugas-tugas pemrosesan bahasa alami (*natural language processing* atau NLP).

## DAFTAR ISI

HALAMAN JUDUL .....	i
HALAMAN PENGESAHAN DOSEN PEMBIMBING .....	ii
HALAMAN PENGESAHAN DOSEN PENGUJI .....	iii
HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR .....	iv
HALAMAN PERSEMBAHAN .....	v
HALAMAN MOTO .....	vi
KATA PENGANTAR .....	vii
SARI .....	ix
GLOSARIUM .....	x
DAFTAR ISI .....	xi
DAFTAR TABEL .....	xiii
DAFTAR GAMBAR .....	xiv
<b>BAB I PENDAHULUAN</b> .....	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	2
1.3 Batasan Masalah .....	3
1.4 Tujuan Penelitian .....	3
1.5 Manfaat Penelitian .....	3
1.6 Sistematika Penulisan .....	3
<b>BAB II LANDASAN TEORI</b> .....	<b>5</b>
2.1 Twitter .....	5
2.2 <i>Text Mining</i> .....	5
2.3 <i>Topic Modeling</i> .....	6
2.4 Metode <i>Topic Modeling</i> .....	6
2.5 Penelitian Terdahulu .....	8
2.6 <i>Latent Dirichlet Allocation (LDA)</i> .....	11
2.7 <i>Preprocessing</i> .....	13
2.7.1 <i>Remove Punctuation</i> .....	14
2.7.2 <i>Case Folding</i> .....	14
2.7.3 <i>Tokenizing</i> .....	14
2.7.4 <i>Stopwords</i> .....	14
2.8 <i>Term Frequency-Invers Document Frequency (TF-IDF)</i> .....	14
2.8.1 <i>Term Frequency (TF)</i> .....	15
2.8.2 <i>Inverse Document Frequency (IDF)</i> .....	15
2.9 Python .....	16
2.10 Anaconda .....	16
2.11 Jupyter Notebook .....	17
2.12 Docker .....	17
<b>BAB III METODOLOGI PENELITIAN</b> .....	<b>18</b>
3.1 Tahapan Penelitian .....	18
3.2 Pengumpulan Data .....	19
3.3 <i>Preprocessing</i> .....	19
3.4 TF-IDF .....	21
3.5 Pemodelan Topik Menggunakan LDA .....	21
3.6 Analisis Hasil .....	22
3.7 Interpretasi Topik .....	22
<b>BAB IV HASIL DAN PEMBAHASAN</b> .....	<b>23</b>

4.1	Pengumpulan Data .....	23
4.2	Implementasi <i>Preprocessing</i> .....	25
4.2.1	<i>Remove Punctuation</i> .....	25
4.2.2	<i>Case Folding</i> .....	26
4.2.3	<i>Tokenizing</i> .....	26
4.2.4	<i>Stopwords</i> .....	27
4.3	Pembobotan.....	28
4.4	Hasil <i>Topic Modeling</i> .....	29
4.5	Iterasi 2.....	43
BAB V PENUTUP .....		56
5.1	Kesimpulan .....	56
5.2	Saran.....	56
DAFTAR PUSTAKA.....		57

**DAFTAR TABEL**

Tabel 2. 1 Perbandingan Metode .....	10
Tabel 4.1 Data hasil <i>scrapping</i> data Twitter .....	25
Tabel 4.2 Contoh <i>tweet</i> setelah dilakukan <i>remove punctuation</i> .....	26
Tabel 4.3 Contoh <i>tweet</i> setelah <i>case folding</i> .....	26
Tabel 4.4 Contoh <i>tweet</i> setelah dilakukan <i>tokenizing</i> .....	27
Tabel 4.5 Contoh <i>tweet</i> setelah dilakukan <i>stopwords</i> .....	28
Tabel 4.6 <i>Coherence Value</i> .....	30
Tabel 4.7 <i>Principal Component (PC)</i> .....	33
Tabel 4.8 Hasil Perbandingan .....	39
Tabel 4.9 <i>Coherence Value</i> .....	45
Tabel 4.10 <i>Principal Component (PC)</i> .....	46
Tabel 4.11 Hasil Perbandingan .....	52

## DAFTAR GAMBAR

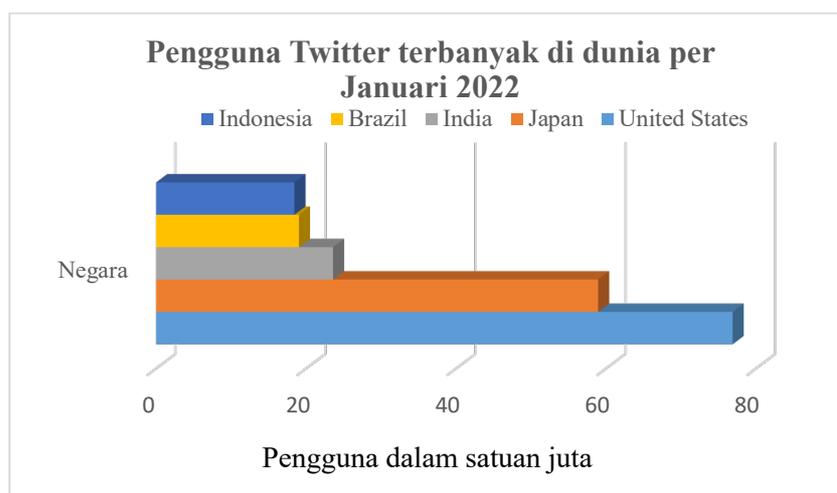
Gambar 2. 1 Macam Metode <i>Topic Modeling</i> .....	7
Gambar 2. 2 Konsep LDA .....	12
Gambar 2. 3 Model Grafis LDA.....	12
Gambar 3. 1 Alur Penelitian. ....	18
Gambar 3. 2 Tahapan <i>Preprocessing</i> .....	20
Gambar 4. 1 Isi Dockerfile.....	23
Gambar 4. 2 Perintah menuju pengambilan data.....	24
Gambar 4. 3 Perintah pengambilan data Twitter. ....	24
Gambar 4. 4 Dataset yang diperoleh.....	24
Gambar 4. 5 Contoh kode program <i>remove punctuation</i> dan <i>case folding</i> .....	26
Gambar 4. 6 Kode Program <i>Tokenizing</i> .....	27
Gambar 4. 7 Kode Program <i>Stopwords</i> .....	28
Gambar 4. 8 Implementasi TF-IDF .....	28
Gambar 4. 9 implementasi untuk menghitung <i>coherence score</i> .....	30
Gambar 4. 10 Grafik <i>Coherence Score</i> kueri “penipuan” .....	30
Gambar 4. 11 Implementasi Visualisasi model LDA dari Tweet “penipuan”.....	31
Gambar 4. 12 Visualisasi LDA dari Tweet “penipuan” .....	32
Gambar 4. 13 Visualisasi <i>Topic-1</i> .....	33
Gambar 4. 14 Visualisasi <i>Topic-2</i> .....	34
Gambar 4. 15 Visualisasi <i>Topic-3</i> .....	34
Gambar 4. 16 Visualisasi <i>Topic-4</i> .....	35
Gambar 4. 17 Visualisasi <i>Topic-5</i> .....	35
Gambar 4. 18 Visualisasi <i>Topic-6</i> .....	36
Gambar 4. 19 Visualisasi <i>Topic-7</i> .....	36
Gambar 4. 20 Visualisasi <i>Topic-8</i> .....	37
Gambar 4. 21 Visualisasi <i>Topic-9</i> .....	37
Gambar 4. 22 Visualisasi <i>Topic-10</i> .....	38
Gambar 4. 23 Dataset terbaru .....	43
Gambar 4. 24 Implementasi Menggabungkan dua file.....	44
Gambar 4. 25 Dataset Gabungan .....	44
Gambar 4. 26 visualisasi grafik <i>Coherence Score</i> .....	45
Gambar 4. 27 Visualisasi LDA dari Tweet “penipuan” .....	46

Gambar 4. 28 Visualisasi <i>Topic-1</i> .....	47
Gambar 4. 29 Visualisasi <i>Topic-2</i> .....	48
Gambar 4. 30 Visualisasi <i>Topic-3</i> .....	48
Gambar 4. 31 Visualisasi <i>Topic-4</i> .....	49
Gambar 4. 32 Visualisasi <i>Topic-5</i> .....	49
Gambar 4. 33 Visualisasi <i>Topic-6</i> .....	50
Gambar 4. 34 Visualisasi <i>Topic-7</i> .....	50
Gambar 4. 35 Visualisasi <i>Topic-8</i> .....	51
Gambar 4. 36 Visualisasi <i>Topic-9</i> .....	51

## BAB I PENDAHULUAN

### 1.1 Latar Belakang

Tidak dapat dimungkiri media sosial merupakan hal yang umum digunakan oleh masyarakat Indonesia. Hal ini sebagaimana yang tertulis dalam “OOSGA” sebuah *global supply chain intelligence platform* yang sering melakukan survei. OOSGA melakukan survei terhadap pengguna media sosial aktif di Indonesia sepanjang tahun 2022. Tercatat sekitar 188,6 juta pengguna media sosial aktif dengan tingkat pertumbuhan tahunan sebesar 12,6% pada tahun 2021 hingga 2022. Hal tersebut mewakili 68,9% dari total populasi yang ada di Indonesia. Pengguna media sosial ini menghabiskan rata-rata waktu 3,28 jam di media sosial (OOSGA, 2023).



Gambar 1.1 Grafik Pengguna Twitter di dunia per Januari 2022

Salah satu media sosial yang banyak digunakan oleh masyarakat Indonesia salah satunya adalah Twitter. Indonesia merupakan salah satu negara dengan pengguna Twitter terbanyak di dunia. Hal tersebut terbukti dengan didudukinya peringkat ke-5 oleh Indonesia, setelah negara Amerika, Jepang, India, dan Brazil. Jumlah negara dengan pengguna Twitter terbanyak di dunia ditunjukkan pada Gambar 1.1. Data tersebut berdasarkan dari laporan platform “Statista” sebuah platform yang berspesialisasi dalam data pasar dan konsumen. Tercatat bahwa jumlah pengguna Twitter di Indonesia per Januari 2022 yaitu sebanyak 18,45 juta, hal tersebut setara dengan 4,32% pengguna Twitter di dunia (Dixon, 2022).

Twitter merupakan media sosial berbasis utama teks. Pada Twitter orang dapat bertukar informasi, berinteraksi dengan orang lain, dan mengikuti perkembangan terbaru dari berbagai topik yang mereka minati. Peruntukan lain dari Twitter yaitu menyediakan ruang bagi pengguna untuk berbagi cerita dan pengalaman mereka, termasuk kasus-kasus penipuan yang mereka alami (Trinugraheni, 2022), (Shalihah, 2021), dan (Khatulistiwa, 2021). Kasus penipuan di Twitter telah menjadi perhatian serius karena dampak negatif yang ditimbulkannya. Individu yang menjadi korban penipuan sering mengalami kerugian finansial, kehilangan informasi pribadi, atau mengalami kerugian lainnya (Yuwita et al., 2022). Oleh karena itu, penting untuk memahami dan mengidentifikasi pola dan karakteristik kasus penipuan yang disebar di Twitter agar dapat mengembangkan strategi yang efektif untuk pencegahan dan perlindungan.

Jumlah informasi mengenai penipuan yang terdapat di Twitter menjadi semakin melimpah, sehingga diperlukan pendekatan yang lebih terstruktur untuk menganalisis dan memahami fenomena tersebut. Sebagai salah satu solusi, dilakukan pemodelan topik terhadap kasus-kasus penipuan yang terjadi di Twitter dalam kurun waktu tiga tahun terakhir, yakni tahun 2019 hingga 2022 sebagai awalan penelitian. Dengan menggunakan metode *Latent Dirichlet Allocation* (LDA), dapat mengklasifikasikan dan mengidentifikasi topik-topik penipuan yang paling umum dibicarakan oleh pengguna Twitter di Indonesia. Pemahaman yang lebih komprehensif tentang topik-topik penipuan ini dapat membantu mengantisipasi dan mencegah kasus-kasus penipuan yang lebih kompleks dan berpotensi merugikan banyak pihak di masa depan. Selain itu, hasil pemodelan topik ini dapat digunakan sebagai dasar untuk memperkuat upaya-upaya pencegahan penipuan yang telah dibicarakan di media sosial dan mengedukasi pengguna Twitter tentang cara-cara untuk mengidentifikasi penipuan yang muncul di platform tersebut.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalahnya yaitu:

- a. Bagaimana mengimplementasikan metode *Latent Dirichlet Allocation* (LDA) pada penelitian yang dilakukan?
- b. Topik apa saja yang dimodelkan dari data Twitter selama kurun waktu 2019 hingga 2022?

### 1.3 Batasan Masalah

Batasan masalah pada penelitian ini yaitu:

- a. Data diambil dari cuitan atau *tweet* di Twitter dalam kurun waktu tiga tahun terakhir, yakni tahun 2019 hingga 2022 dengan menggunakan *keyword* “penipuan”.
- b. Data cuitan atau *tweet* yang diambil menggunakan bahasa Indonesia.
- c. Data yang diambil terbatas pada wilayah Indonesia.

### 1.4 Tujuan Penelitian

Tujuan dari penelitian ini yaitu:

- a. Memperoleh gambaran umum mengenai *tweet* terkait penipuan yang terjadi di Indonesia pada media sosial Twitter.
- b. Dapat mengklasterisasi topik informasi penipuan apa saja yang sering dibahas di Twitter dalam kurun waktu tiga tahun terakhir, yakni tahun 2019 hingga 2022.

### 1.5 Manfaat Penelitian

Manfaat dari penelitian ini yaitu:

- a. Memberikan pengetahuan kepada pembaca mengenai *topic modeling* menggunakan metode LDA terkait informasi penipuan yang tersebar di media sosial twitter di Indonesia.
- b. Agar pembaca dapat lebih waspada terhadap jenis-jenis penipuan yang terjadi yang sering dibahas di Twitter.

### 1.6 Sistematika Penulisan

Sistematika penulisan dalam penyusunan tugas akhir ini terdiri dari beberapa bab, yang mencakup gambaran dari keseluruhan masalah dan penyelesaiannya. Berikut sistematika penulisan yang terbagi dalam 5 bab:

#### BAB I PENDAHULUAN

Pada bab ini berisi pembahasan latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

#### BAB II LANDASAN TEORI

Pada bab ini berisi pembahasan mengenai konsep dan prinsip dasar untuk memecahkan masalah yang ada di penelitian. Terdapat juga uraian tentang hasil penelitian yang pernah

dilakukan sebelumnya oleh peneliti lain yang memiliki kemiripan dengan penelitian yang dilakukan.

### BAB III METODOLOGI PENELITIAN

Bab ini berisi tentang tahapan penelitian yang akan dilakukan seperti pengambilan data, *preprocessing*, TF-IDF, pemodelan topik menggunakan LDA, analisis hasil, dan interpretasi topik.

### BAB IV HASIL DAN PEMBAHASAN

Bab ini akan membahas tentang pengimplementasian metode dan hasilnya.

### BAB V KESIMPULAN DAN SARAN

Bab ini merupakan bab terakhir yang akan membahas kesimpulan dan saran terhadap penelitian yang telah dilakukan pada tugas akhir.

## **BAB II**

### **LANDASAN TEORI**

#### **2.1 Twitter**

Twitter merupakan salah satu media sosial yang digunakan untuk berkomunikasi dalam pesan singkat yang sering disebut cuitan atau *tweets* (Kompas.com, 2022). Twitter menggunakan gambar burung biru sebagai logo untuk melambangkan pesan-pesan atau *tweets* yang terbang dan tersebar di seluruh dunia dengan cepat, seperti burung yang terbang dengan cepat dan mengeluarkan cuitan (Firdausi, 2022). Cuitan “*just setting up my twttr*” adalah cuitan/*tweet* pertama yang diluncurkan pada 22 Maret 2006 oleh pendiri Twitter, Jack Dorsey yang membuat tanggal tersebut ditetapkan menjadi tanggal berdirinya Twitter (Kompas.com, 2021).

Menurut Shayne (2021), jenis media sosial Twitter yaitu “*Social Network*” terlihat dari fokus Twitter dalam membantu penggunanya terhubung dengan banyak orang, membangun komunitas maupun grup, serta sebagai wadah penggunanya untuk berbagi ide, minat, dan informasi yang mereka miliki (Shayne, 2021). Santoni & Rufat (2021), menggunakan media sosial Twitter untuk menguji seberapa cepat pengguna Twitter dalam melaporkan kejadian darurat. Tujuannya adalah untuk mengevaluasi apakah Twitter dapat digunakan sebagai alat manajemen darurat (SMEM) melalui analisis yang dilakukan. Dengan mengandalkan kebiasaan pengguna Twitter yang aktif untuk memperbarui informasi yang diterima. Informasi tersebut bisa berupa pengalaman langsung pengguna Twitter atau informasi yang disebarluaskan oleh teman-teman pengguna Twitter. Menurut Polyzos & Wang, (2022), Twitter dapat dianggap sebagai representasi yang sangat dekat dari penyebaran informasi yang tersedia secara publik. Pengguna umumnya membagikan informasi yang sudah diketahui (atau akan diketahui oleh sebagian besar orang setelah *tweet* diposting). Hal tersebut membuktikan jika media sosial Twitter mampu menjadi sumber data untuk melakukan sebuah penelitian yang membutuhkan informasi yang terbaru.

#### **2.2 Text Mining**

*Text Mining* merupakan proses pengambilan informasi dari dokumen atau teks tidak terstruktur menggunakan teknik-teknik komputasi seperti pembelajaran mesin atau *Machine Learning* (ML), statistik, dan pemrosesan bahasa alami atau *Natural Language Processing*

(NLP). Tujuan dari *text mining* sendiri yaitu mengekstrak informasi yang berguna dari teks dan mengubahnya menjadi pengetahuan yang dapat digunakan untuk pengambilan keputusan. Proses tersebut melibatkan teknik-teknik seperti tokenisasi, *stemming*, dan penghapusan *stopword* untuk mengurangi kompleksitas data dan membuatnya lebih mudah untuk dianalisis. Setelah itu, teknik-teknik seperti klasifikasi, korelasi, dan klusterisasi dapat digunakan untuk mengolah data dan menghasilkan informasi yang bermanfaat (Bach et al., 2019).

### **2.3 Topic Modeling**

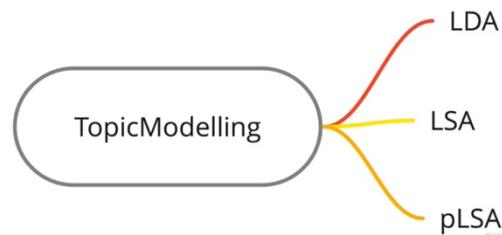
*Topic modeling* merupakan salah satu teknik analisis data yang digunakan untuk mengelompokkan dokumen teks ke dalam topik atau tema tertentu. *Topic modeling* termasuk teknik *text mining* yang paling umum digunakan dalam pemrosesan bahasa alami atau *Natural Language Processing* (NLP) untuk mendapatkan pemahaman yang lebih mendalam tentang isi dokumen teks (Manthiramoorthi, 2018). Algoritma dalam *topic modeling* digunakan untuk mengelompokkan kata-kata ke dalam kelompok topik yang berbeda dan menentukan tingkat keanggotaan setiap dokumen dalam setiap topik. *Topic modeling* termasuk metode *soft* atau *fuzzy clustering* dalam *unsupervised learning*. Dalam metode *soft clustering*, setiap dokumen dapat dimiliki oleh lebih dari satu topik dengan tingkat keanggotaan yang berbeda-beda. Sehingga, meskipun suatu dokumen cenderung masuk ke dalam satu topik tertentu, namun dokumen tersebut masih dapat dimiliki oleh topik lainnya dengan tingkat keanggotaan yang lebih rendah (Arianto & Anuraga, 2020).

Tujuan utama dari *topic modeling* adalah untuk membantu dalam mengidentifikasi pola dan tren yang tersembunyi dalam rangkaian kata pada kumpulan dokumen yang tidak terstruktur. Dengan menggunakan *topic modeling*, memudahkan untuk memahami topik utama yang dibahas dalam dokumen tersebut, serta hubungan antara topik atau tema yang berbeda di antara kumpulan dokumen. Selain itu, *topic modeling* juga dapat membantu dalam melakukan analisis teks, pengelompokkan dokumen, dan penyederhanaan data, sehingga memudahkan pemrosesan data dan pengambilan keputusan (Chilmi, 2021).

### **2.4 Metode Topic Modeling**

Pada pengaplikasian *topic modeling*, dibutuhkan metode yang membantu jalannya proses tersebut. Penggunaan metode dapat membantu meningkatkan kualitas hasil *topic modeling* sehingga dapat digunakan untuk analisis atau pengambilan keputusan yang lebih baik. Menurut

Mandal (2020), Gambar 2.1 menunjukkan tiga jenis metode *topic modeling* yang umum digunakan dalam penelitian.



Gambar 2. 1 Macam Metode *Topic Modeling*

a. *Latent Dirichlet Allocation (LDA)*

LDA merupakan model probabilistik generatif yang digunakan untuk memproses data diskret seperti teks. Model ini merepresentasikan dokumen sebagai campuran acak dari topik laten (tidak terlihat). LDA merupakan model Bayesian hirarki tiga tingkat, di mana setiap item koleksi dimodelkan sebagai campuran terbatas dari serangkaian set topik, dan setiap topik dimodelkan sebagai campuran tak terbatas melalui set yang mendasari probabilitas topik. Dalam konteks pembuatan model teks, probabilitas topik memberikan representasi eksplisit dari sebuah dokumen (Suhartono, 2018).

LDA bekerja dengan memodelkan setiap dokumen sebagai kombinasi dari beberapa topik yang tidak terlihat (disebut juga sebagai distribusi topik) dan setiap topik sebagai distribusi probabilitas atas kata-kata yang mungkin muncul dalam topik tersebut. Tujuan dari LDA untuk menemukan distribusi topik yang paling mungkin di balik sebuah dokumen, dan distribusi kata-kata yang paling mungkin di balik setiap topik (ER, 2021).

b. *Latent Semantic Analysis (LSA)*

Tujuan dari teknik analisis teks yang disebut *Latent Semantic Analysis (LSA)* adalah untuk menggunakan konteks kata-kata dalam sebuah dokumen untuk mengungkap ide atau subjek tersembunyi di dalamnya. Dalam pendekatan ini, komputer melakukan analisis pada dokumen menggunakan metode *Term Frequency-Inverse Document Frequency (TF-IDF)*. Statistik TF-IDF yang digunakan dalam teknik LSA ini dapat menghitung tingkat signifikansi sebuah kata dalam kumpulan dokumen, sehingga memungkinkan untuk mengidentifikasi kata-kata kunci yang paling penting dalam dokumen tersebut (Jayadianti et al., 2020).

c. *Probabilistic Latent Semantic Analysis (pLSA)*

Merupakan sebuah model probabilistik yang digunakan untuk mengganti metode *Singular Value Decomposition* (SVD) dalam *Latent Semantic Analysis* (LSA). PLSA membantu dalam mengatasi beberapa tantangan yang terkait dengan representasi dalam LSA. LSA menggunakan matriks *Term-Frequency Inverse-Document-Frequency* (TF-IDF) untuk merepresentasikan dokumen dan *term* dalam bentuk numerik. Representasi ini kemudian diolah menggunakan metode SVD untuk mengurangi dimensi dan memperoleh representasi semantik yang lebih baik. Namun, SVD tidak memiliki dasar probabilitas yang jelas, yang menyebabkan beberapa masalah (Goyal, 2021).

PLSA mengatasi masalah ini dengan merepresentasikan setiap entri dalam matriks TF-IDF menggunakan probabilitas. Model probabilistik ini mengasumsikan bahwa setiap dokumen dapat dilihat sebagai distribusi probabilitas atas topik-topik tertentu, dan setiap topik dapat dilihat sebagai distribusi probabilitas atas *term-term* tertentu. Oleh karena itu, pLSA dapat digunakan untuk menemukan topik-topik yang paling relevan dalam sebuah dokumen atau koleksi dokumen (Goyal, 2021).

## 2.5 Penelitian Terdahulu

Dalam pembahasan ini mengambil 6 penelitian *topic modeling* terdahulu sebagai perbandingan dari setiap metode pengembangan sistem.

Pada penelitian yang berjudul "*Topic Modeling* Penelitian Dosen JPTEI UNY pada Google Scholar Menggunakan *Latent Dirichlet Allocation*" yang dilakukan oleh Nurlayli & Nasichuddin (2019), menghadirkan analisis tentang pemodelan topik penelitian berdasarkan data judul publikasi beberapa dosen dari Jurusan Pendidikan Teknik Elektro dan Informatika Industri (JPTEI) di Universitas Negeri Yogyakarta. Data tersebut diambil dari Google Scholar untuk mengetahui kecenderungan penelitian yang dilakukan oleh para dosen pada suatu jurusan atau program studi tertentu. Dalam penelitian ini, metode *Latent Dirichlet Allocation* (LDA) digunakan untuk membantu mengidentifikasi topik-topik penelitian yang sering dibahas oleh para dosen di JPTEI UNY.

Selanjutnya, Polyzos & Wang (2022) melaksanakan penelitian dengan judul "*Twitter and Market Efficiency in Energy Markets: Evidence Using LDA Clustered Topic Extraction*". Penelitian ini bertujuan untuk menguji dan mengukur efisiensi pasar energi dengan menggunakan sampel data *tweet*. Mereka melakukan pengumpulan topik-topik yang muncul pada hari-hari di mana indeks pasar energi naik dan turun menggunakan algoritma *Latent*

*Dirichlet Allocation* (LDA). Hasil penelitian menunjukkan adanya perbedaan topik-topik *tweet* pada hari-hari ketika indeks pasar energi mengalami kenaikan dan penurunan.

Negara et al., (2019) juga melakukan sebuah penelitian yang berjudul "*Topic Modeling Twitter with Latent Dirichlet Allocation Method*". Tujuan penelitian ini adalah untuk mengetahui topik-topik yang sering dibahas di media sosial Twitter dalam empat kategori topik yang telah ditentukan, yaitu "*Economic, Military, Sports, Technology*". Penulis juga membandingkan dua metode, yaitu LDA dan LSI, untuk melakukan *Topic Modeling* guna mencapai hasil yang optimal.

Selanjutnya, Rashif et al., (2021) melaksanakan penelitian dengan judul "*Implementasi LDA untuk Pengelompokan Topik Cuitan Akun Bot Twitter bertagar #Covid-19 LDA*". Penelitian ini bertujuan untuk mengidentifikasi berbagai topik yang dibahas oleh akun bot terkait penyebaran informasi dengan tagar #Covid19 di Twitter. Dalam penelitian ini, metode LDA digunakan sebagai alat analisis utama.

Pada Hassan et al., (2020) melaksanakan penelitian berjudul "*LSA & LDA Topic Modeling Classification: Comparison study on E-books*". Penelitian ini membandingkan efektivitas dari dua pendekatan pemodelan topik yang populer, yaitu LDA dan LSA, dalam mengklasifikasikan dokumen teks ilmiah yang tidak terstruktur, seperti *e-book*. Penelitian ini dilakukan sebagai respons terhadap pertumbuhan pesat teknologi informasi yang menyebabkan meningkatnya jumlah data teks yang tidak terstruktur di perpustakaan digital. Oleh karena itu, dibutuhkan penelitian untuk menganalisis, mengorganisasi, dan mengklasifikasikan teks secara otomatis di repositori penelitian *online*. Pemodelan topik digunakan sebagai alat untuk mengklasifikasikan dokumen secara otomatis, dan penulis juga membandingkan kinerja model LDA dan LSA.

Terakhir, Gupta & Patel (2021) melaksanakan penelitian dengan judul "*Method of Text Summarization Using LSA And Sentence Based Topic Modeling with Bert*". Penelitian ini membahas tugas penting dalam bidang *Natural Language Processing* (NLP), yaitu meringkas dokumen dan pemodelan topik. Tujuan utamanya adalah untuk membuat ringkasan singkat dan padat dari data teks yang panjang, tetapi tetap mengandung informasi yang relevan dalam dokumen. Dalam penelitian ini, metode LSA digunakan sebagai alat untuk mengekstrak topik yang relevan dari dokumen teks.

Tabel 2.1 menunjukkan perbandingan metode yang dimiliki penelitian terdahulu yang telah disebutkan.

Tabel 2.1 Perbandingan Metode

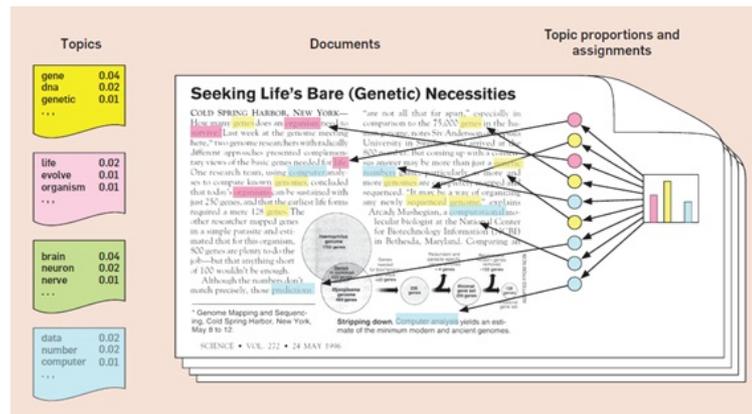
Penulis, Tahun	Metode	Keterangan
1. (Nurlayli & Nasichuddin, 2019) 2. (Polyzos & Wang, 2022) 3. (Negara et al., 2019) 4. (Rashif et al., 2021)	LDA	<ul style="list-style-type: none"> <li>• Tujuan penelitian ini yaitu mengimplementasikan pemodelan topik pada judul publikasi JPTEIUNY di Google Scholar. <i>Latent Dirichlet Allocation</i> (LDA) digunakan sebagai metode. Hasil yang diperoleh bahwa dosen JPTEI-UNY biasanya memiliki judul penelitian tentang pendidikan vokasi, pengembangan sistem, media pembelajaran, dan sistem pembelajaran di SMK.</li> <li>• Penelitian ini menggunakan sampel <i>tweet</i> pasar energi yang untuk mempelajari dan mengukur efisiensi pasar. Menggunakan LDA sebagai metode untuk membantu menguraikan hari-hari kenaikan dan penurunan topik untuk menemukan bahwa topik <i>tweet</i> pada hari-hari kenaikan dan penurunan berbeda satu sama lain. Hasil dari pengklasifikasian menemukan bahwa pengklasifikasi memberikan prakiraan pergerakan pasar dengan akurasi 57,83% (39,02%) di pasar <i>bull</i> (<i>bear</i>).</li> <li>• Penelitian ini membandingkan dua metode (LDA &amp; LSI) sebagai metode yang membantu jalannya proses <i>Topic Modeling</i> terhadap 4 topik (<i>Economic, Military, Sports, Technology</i>) yang sering dibahas di Twitter. Hasil dari penelitian ini menunjukkan menggunakan metode LDA lebih optimal daripada metode LSI terlihat dari akurasi yang didapat untuk topik "<i>sport</i>" dari 1260 <i>tweet</i> yang memiliki akurasi 98%.</li> <li>• Penelitian ini melakukan analisis pada 162 <i>tweet</i> dari 62 akun bot Twitter. Menggunakan metode LDA dan menemukan hasil lima topik teratas antara lain tentang kondisi dan dampak pandemic saat ini, himbauan untuk menjaga jarak agar Kesehatan tetap terjaga, perkembangan penyebaran Covid-19 yang ada di Indonesia, vaksinasi yang terjadi di beberapa wilayah di Indoneisa, dan cara menghadapi Covid-19.</li> </ul>

5. (Hassan et al., 2020)	LSA	<ul style="list-style-type: none"> <li>• Penelitian ini untuk mengklasifikasikan teks secara otomatis dalam repositori <i>e-research</i> dengan menggunakan <i>dataset</i> yang terdiri dari (300) buku berisi sekitar 23 juta kata berdasarkan teks lengkap. Menggunakan model LSA dan LDA. Menunjukkan hasil bahwa LDA dengan nilai koherensi (0.593179) memiliki hasil yang lebih baik dari LSA dengan nilai koherensi (0.55777302) dengan jumlah 10 topik.</li> </ul>
6. (Gupta & Patel, 2021)		<ul style="list-style-type: none"> <li>• Penelitian ini melakukan <i>document summarization</i>. Terdapat tahapan <i>topic modeling</i> menggunakan model LSA menggunakan SVD terpotong yang mengekstrak semua topik yang relevan dari teks.</li> </ul>

Pada penelitian yang menggunakan metode PLSA untuk pemodelan topik jarang digunakan, dikarenakan metode PLSA memiliki kompleksitas komputasi yang berpotensi tinggi, cenderung mengalami *overfitting*, dan bukan model *generative* yang terdefinisi dengan baik untuk dokumen baru (Goyal, 2021). Berdasarkan hasil perbandingan yang dilakukan, penelitian ini akan menggunakan metode *Latent Dirichlet Allocation* (LDA). Metode LDA dinilai tepat dalam meningkatkan kualitas hasil *topic modeling* dikarenakan model ini merupakan model yang paling sesuai dengan penelitian yang akan dilakukan dikarenakan model ini dapat menemukan topik tersembunyi dalam dokumen dan menganalisis data teks.

## 2.6 *Latent Dirichlet Allocation* (LDA)

LDA atau *Latent Dirichlet Allocation* merupakan teknik *unsupervised clustering* yang digunakan untuk analisis teks. LDA merupakan jenis model topik di mana kata-kata direpresentasikan sebagai topik dan dokumen direpresentasikan sebagai koleksi topik kata-kata. Dengan menggunakan LDA, kita dapat menemukan topik tersembunyi dalam dokumen dan menganalisis data teks secara lebih efisien dan efektif (Zvornicanin, 2022).



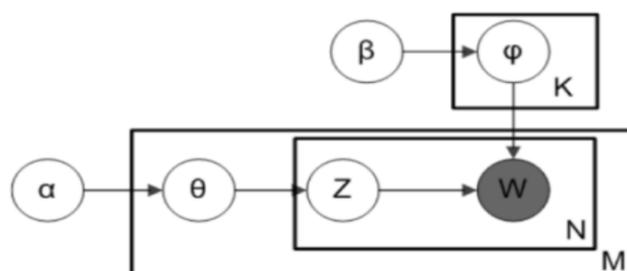
Gambar 2. 2 Konsep LDA

(Sumber: <https://socs.binus.ac.id/2018/11/29/latent-dirichlet-allocation-lda/> )

Cara kerja dari model LDA seperti terlihat pada Gambar 2.2 di mana dapat diasumsikan topik telah dispesifikasikan sebelum didapatkan dokumen pada gambar tersebut merupakan deretan topik yang ada di sebelah kiri.

Untuk setiap dokumen di dalam koleksi dilakukan:

1. Secara acak dipilih distribusi atas topik (pada gambar ditunjukkan sebagai grafik distribusi topik di sebelah kanan).
2. Untuk setiap kata di dalam dokumen:
  - Secara acak dipilih sebuah topik dari distribusi atas topik, pada langkah 1 (pada gambar ditunjukkan pada hubungan grafik dengan lingkaran).
  - Secara acak dipilih distribusi sebuah kata dari distribusi yang sesuai, atas kosakata. (pada gambar ditunjukkan dengan cara memilih warna pada lingkaran)



Gambar 2. 3 Model Grafis LDA.

(Sumber : <https://socs.binus.ac.id/2018/11/29/latent-dirichlet-allocation-lda/> )

LDA digambarkan dengan model grafis menggunakan *plate notation* seperti Gambar 2.4, yang di mana:

- $\beta$  merupakan *dirichlet* parameter atas distribusi kata terhadap topik
- $\phi$  merupakan distribusi kata terhadap topik dalam *corpus*
- $K$  merupakan kumpulan topik
- $W$  merupakan kata
- $N$  merupakan kumpulan kata
- $M$  merupakan kumpulan dokumen
- $Z$  merupakan topik *index assignment*
- $\theta$  merupakan dokumen, dan
- $\alpha$  merupakan *dirichlet* parameter atas distribusi topik terhadap dokumen

(Zvornicanin, 2022)

LDA dirumuskan sebagai berikut:

$$p(w,z,\theta,\phi|\alpha,\beta) = p(\phi|\beta)p(\theta|\alpha)p(z|\theta)p(w|\phi) \quad (2.1)$$

Pada implementasinya LDA tidak dapat diterapkan karena variabel  $Z$ ,  $\theta$ , dan  $\phi$  yang tersembunyi atau tidak diketahui dan juga untuk mencari asosiasi antara  $z$  dan  $w$  sulit karena satu kata dapat mengandung 2 topik atau lebih, dan ketika mencoba untuk mengestimasi  $p(Z|W)$  dapat menghasilkan:

$$p(\vec{z}|\vec{w}) = \frac{p(\vec{z},\vec{w})}{p(\vec{w})} = \frac{\prod_{i=1}^w p(z_i, w_i)}{\prod_{i=1}^w \sum_{k=1}^K p(z_i = k, w_i)} \quad (2.2)$$

## 2.7 Preprocessing

*Preprocessing* merupakan suatu proses awal sebelum melakukan analisis data agar data yang akan diolah menjadi terorganisir. Tujuan dari tahapan ini yaitu menghilangkan atau menghapus data yang tidak relevan sehingga proses selanjutnya akan berjalan lebih efektif dan efisien. *Preprocessing* memiliki beberapa tahapan di antaranya yaitu *remove punctuation*, *case folding*, *tokenizing*, dan *stopwords* (Chilmi, 2021).

### 2.7.1 *Remove Punctuation*

Tahap pertama yang dilakukan yaitu *remove punctuation* yang merupakan proses menghilangkan karakter yang tidak diperlukan seperti angka, tanda baca, *link/tag*, karakter special (% , \$ , & , etc). Tujuan dari dilakukannya *remove punctuation* yaitu untuk membersihkan teks dari elemen-elemen yang tidak relevan atau tidak diperlukan sehingga mempermudah analisis atau pemrosesan selanjutnya (Chilmi, 2021).

### 2.7.2 *Case Folding*

Tahap selanjutnya yaitu *case folding*, tahapan ini akan mengubah karakter dari huruf yang ada dalam data. Karakter akan diubah dari huruf besar menjadi huruf kecil. Tujuan dari tahap ini untuk menghindari pembacaan kata yang sama tetapi dianggap berbeda karena perbedaan huruf besar dan kecil (Chilmi, 2021).

### 2.7.3 *Tokenizing*

Tahap terakhir dari *preprocessing* yaitu *tokenizing*. *Tokenizing* atau tokenisasi merupakan proses memisahkan teks menjadi unit-unit yang lebih kecil disebut “token”. Tujuannya yaitu memudahkan proses perhitungan kata hingga transformasi data (Chilmi, 2021).

### 2.7.4 *Stopwords*

Tahap selanjutnya yang dilakukan yaitu *stopwords*. Dalam tahap ini, akan dibuat *stoplist* yang berisikan daftar kata-kata yang tidak relevan atau tidak penting. Tujuan tahap ini untuk memproses kata-kata tidak penting dan kata-kata yang tersisa akan dianggap sebagai kata-kata penting atau *keywords* (Chilmi, 2021).

## 2.8 *Term Frequency-Invers Document Frequency (TF-IDF)*

*Term Frequency-Invers Document Frequency* atau TF-IDF merupakan sebuah metode yang digunakan untuk menghitung frekuensi sebuah kata di dalam suatu artikel, dokumen, dan frekuensi dari banyak dokumen. Tujuan dari TF-IDF untuk menentukan bobot suatu kata di dalam satu atau banyak dokumen (Sasmita & Falani, 2018).

Rumus TF-IDF akan dijelaskan sebagai berikut (Sasmita & Falani, 2018) :

$$tfidf_{t,d} = tf_{t,d} \cdot idf_t \quad (2.3)$$

Keterangan:

$tfidf_{t,d}$  = bobot kata

$tf_{t,d}$  = *term frequency*

$idf_t$  = *inverse document frequency*

### 2.8.1 Term Frequency (TF)

*Term Frequency* (TF) merupakan jumlah munculnya sebuah kata di dalam dokumen.

Rumus dari TF ditunjukkan pada persamaan (2.4):

$$tf_{t,d} = \begin{cases} \log f_{t,d} + 1 & \text{Jika } f_{t,d} > 0 \\ 0 & \text{Jika kata tidak muncul} \end{cases} \quad (2.4)$$

Keterangan

t = jumlah kata

d = dokumen

$tf_{t,d}$  = *Term Frequency*

$f_{t,d}$  = total kata yang ada dalam dokumen

(Sasmita & Falani, 2018)

### 2.8.2 Inverse Document Frequency (IDF)

*Inverse Document Frequency* merupakan nilai untuk menentukan seberapa penting sebuah kata. Kata yang sering muncul akan dianggap sebagai kata yang kurang penting oleh IDF. Rumus dari IDF yaitu:

$$idf_t = \log \frac{N}{N_t} \quad (2.5)$$

Keterangan

$idf_t$  = *Inverse Document Frequency*

N = Jumlah Dokumen

$N_t$  = Jumlah dokumen yang berisi kata t

(Sasmita & Falani, 2018)

## 2.9 Coherence Score

*Coherence Score* merupakan metrik yang digunakan untuk menilai seberapa baik sebuah model dalam menghasilkan topik-topik yang terkait, di mana model yang unggul akan menghasilkan topik dengan *coherence score* yang tinggi (Narendra, 2022). Nilai *coherence score* lebih dari nol menandakan bahwa topik tersebut layak dipertimbangkan untuk digunakan (Zvornicanin, 2023).

## 2.10 Python

Perusahaan besar dan pengembang sering menggunakan Python, bahasa pemrograman, untuk membuat berbagai aplikasi desktop, online, dan seluler. Nama Python, yang berasal dari Sirkus Terbang Monty Python, ditemukan oleh Guido van Rossum di Belanda pada tahun 1990. Python diciptakan oleh Van Rossum sebagai hobi, dan sebagai hasil dari sintaksisnya yang mudah dipahami dan pustaka yang besar, sekarang telah mendapatkan popularitas baik di industri maupun pendidikan (Clinton & Sengkey, 2019).

Penggunaan Python dalam *machine learning* sering dipilih dikarenakan Python merupakan bahasa pemrograman yang mudah dipelajari. Python juga menyediakan banyak *library* yang memudahkan programmer *data scientist* untuk menyelesaikan permasalahan yang rumit dengan waktu yang singkat. Contoh *library* yang disediakan oleh Python di antaranya Tensorflow untuk *machine learning*, spaCy untuk *natural language processing*, NumPy untuk *computational science*, OpenCV untuk *computer vision*, dan masih banyak lagi (Nursyafitri, 2022).

## 2.11 Anaconda

Anaconda merupakan sebuah platform bahasa pemrograman Python dan R didistribusikan bersama dalam sebuah paket yang disebut Anaconda, yang juga mencakup sejumlah paket lain untuk pemrograman ilmu data dan komputasi ilmiah, termasuk untuk *data science*, *machine learning*, *data processing* skala besar, analisis prediktif, matematika, dan rekayasa. Banyak pustaka dan paket *pra-instal* tersedia dengan Anaconda. NumPy, SciPy, Pandas, Scikit-Learn, nltk, dan Jupiter merupakan beberapa di antaranya. Sementara Python menawarkan pip sebagai manajer paket, Anaconda menawarkan conda sebagai alternatif pengelolaan paket. Ketergantungan Python dapat diinstal dengan Pip. Penginstallan dependensi library Python dan non-Python dimungkinkan dengan Anaconda (Ulfa, 2022).

### **2.12 Jupyter Notebook**

Jupyter Notebook merupakan aplikasi web sumber terbuka yang memungkinkan untuk membuat dan berbagi dokumen yang berisi kode langsung, persamaan, visualisasi, dan teks naratif. Penggunaannya meliputi *data cleaning and transformation*, *numerical simulation*, *statistical modelling*, *data visualization*, *machine learning*, dan masih banyak lagi (Pryke, 2020).

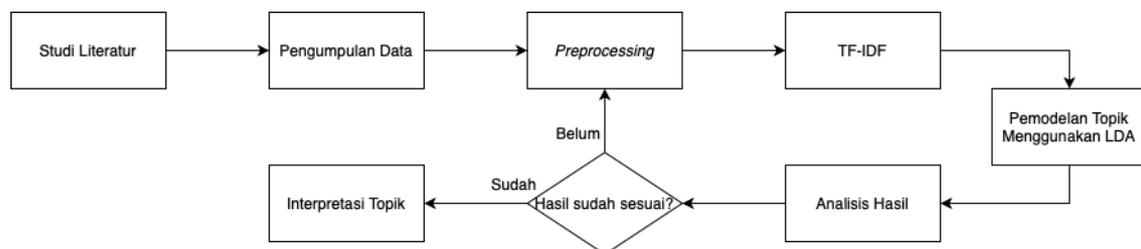
### **2.13 Docker**

Docker yaitu sebuah aplikasi atau platform yang digunakan untuk menggabungkan berkas perangkat lunak yang akan diuji ke dalam sebuah wadah atau *container*. Dengan demikian, proses pengembangan menjadi lebih mudah dilakukan karena semua berkas yang diperlukan sudah terkumpul dalam satu wadah atau container tersebut (Yuliana W, 2023).

## BAB III METODOLOGI PENELITIAN

### 3.1 Tahapan Penelitian

Tahapan penelitian yang akan dilakukan memiliki tujuh tahapan, sebagaimana yang ditunjukkan pada Gambar 3.1.



Gambar 3.1 Alur Penelitian

Tahapan penelitian pertama dimulai dengan melakukan studi literatur mengenai *topic modeling* yang dilakukan oleh penelitian terdahulu lalu membandingkan tiap metode yang digunakan sebagai pertimbangan metode mana yang akan digunakan. Selanjutnya dilakukan proses pengumpulan data *tweet* dari Twitter dalam format *csv* yang digunakan sebagai data yang akan diolah. Setelah data *tweet* berhasil terkumpul akan dilakukan *preprocessing* yang berisi *cleaning* dan *pra-process* data, termasuk menghilangkan karakter yang tidak diperlukan, mengubah huruf menjadi huruf kecil, menghapus kata-kata umum (*stopwords*), tokenisasi, mengubah teks menjadi matriks frekuensi kata atau frekuensi kata-dokumen terbalik (TF-IDF). Selanjutnya setelah data dibersihkan melalui proses *preprocessing*, dilakukan pembobotan menggunakan TF-IDF yang digunakan untuk mengukur pentingnya suatu kata dalam dokumen atau korpus berdasarkan frekuensi kata dalam dokumen tersebut (TF) dan seberapa umum kata tersebut dalam seluruh korpus (IDF). Setelah menemukan bobot kata, lalu dilakukannya pemodelan topik menggunakan LDA. LDA merupakan sebuah model probabilistik yang digunakan untuk mengidentifikasi topik-topik yang tersembunyi dalam sebuah korpus

dokumen. Setelah melakukan pemodelan topik menggunakan LDA, dilakukan analisis hasil untuk mengetahui apakah kata yang didapat dari model tersebut sesuai dengan hasil yang diinginkan. Jika tidak, akan kembali ke proses *preprocessing* untuk membersihkan kata yang tidak perlu. Kemudian, pada tahap terakhir, jika model telah sesuai, akan dilakukan interpretasi topik dengan mengevaluasi kata-kata teratas yang terkait dengan setiap topik dan mencoba memahami topik-topik tersebut. Hal ini penting untuk memperoleh pemahaman yang lebih baik tentang konten dan makna yang terkandung dalam setiap topik.

### 3.2 Pengumpulan Data

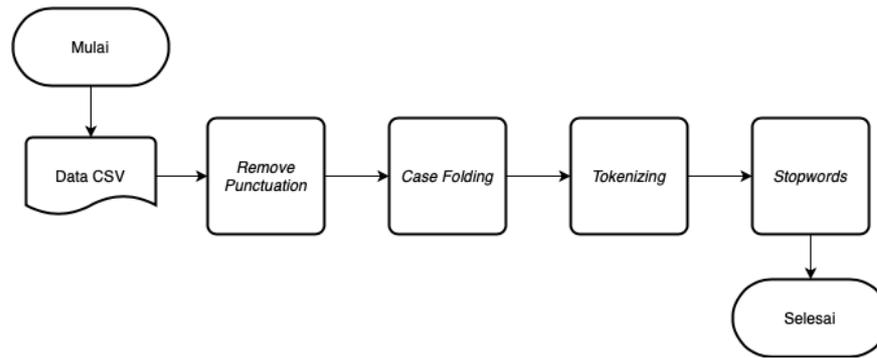
Pengumpulan data dilakukan dengan cara *scraping* Twitter menggunakan Twint yang merupakan sebuah *library* yang digunakan untuk melakukan *data scraping* dari Twitter, yang dapat dijalankan melalui bahasa pemrograman Python. Dalam ekosistem Python, Twint merupakan salah satu *library* populer yang digunakan untuk tujuan *data scraping* Twitter. Twint menyediakan antarmuka baris perintah (*command-line interface*) yang dapat diakses melalui terminal, dan juga dapat digunakan sebagai modul dalam program Python. Twint dapat digunakan untuk mengambil berbagai jenis data Twitter, seperti *tweet*, metadata pengguna, gambar, dan video, dengan beberapa pilihan filter dan opsi pengaturan. Untuk melakukan pengumpulan data, dibutuhkan *software* yang mendukung seperti:

- a. Python 3.9
- b. Docker
- c. Twint *library*
- d. Terminal

Untuk penelitian ini data diambil dalam kurun waktu tiga tahun terakhir, yakni tahun 2019 hingga 2022 dengan kata kunci “penipuan” sebagai *query*. Pada tahap pengambilan data, dialankan perintah di terminal dengan bantuan *software* Docker untuk meminimalisir adanya kesalahan ketika melakukan *data scraping*. Lalu data akan tersimpan dalam bentuk csv pada folder yang sudah ditetapkan sebelumnya.

### 3.3 Preprocessing

*Preprocessing* data merupakan langkah penting dalam analisis teks yang bertujuan untuk membersihkan dan mempersiapkan data teks sebelum diolah lebih lanjut. Beberapa langkah yang dapat dilakukan dalam *preprocessing* dapat dilihat pada Gambar 3.2.



Gambar 3.2 Tahapan *Preprocessing*

*Preprocessing* dilakukan untuk mencegah kehilangan nilai data, kelebihan data, dan ketidakkonsistenan data. Pada penelitian ini, beberapa tahap *preprocessing* yang akan dilakukan seperti:

a. *Remove Punctuation*

Proses menghapus tanda baca dari teks. Tanda baca seperti koma, titik, tanda tanya, tanda seru, dan lainnya dianggap tidak memberikan informasi penting dalam analisis teks, sehingga dihapus agar tidak mempengaruhi hasil analisis.

b. *Case Folding*

*Case Folding* merupakan proses mengubah seluruh huruf dalam teks menjadi huruf kecil atau huruf besar, sehingga perbedaan kapitalisasi tidak lagi menjadi masalah dalam analisis teks. Dalam beberapa kasus, istilah ini juga disebut sebagai *case normalization*.

c. *Tokenizing*

Proses memisahkan teks menjadi token atau bagian-bagian yang lebih kecil seperti kata, frasa, atau kalimat. Tujuan dari *tokenizing* untuk mempermudah pengolahan teks lebih lanjut, seperti penerapan analisis statistik atau pemodelan bahasa.

d. *Stopwords*

*Stopwords* sendiri merupakan kata-kata yang umum digunakan dalam bahasa tertentu dan tidak memiliki makna khusus seperti "ada", "di", "yang", "pada", "untuk", "aku", "kamu", dan kata lainnya. *Stopwords* biasanya dihapus dari teks karena tidak membawa informasi yang signifikan dalam analisis teks.

### 3.4 TF-IDF

TF-IDF (*Term Frequency-Inverse Document Frequency*) merupakan proses dalam pengolahan teks yang digunakan untuk mengevaluasi pentingnya sebuah kata kunci dalam sebuah dokumen atau korpus teks. Metode ini sering digunakan dalam analisis teks dan pemrosesan bahasa alami. Proses TF-IDF melibatkan dua konsep utama yaitu frekuensi kata (*term frequency*) dan frekuensi dokumen terbalik (*inverse document frequency*).

*Term Frequency* (TF) mengukur seberapa sering sebuah kata kunci muncul dalam sebuah dokumen. Tujuannya untuk memberikan bobot yang lebih tinggi pada kata-kata yang muncul lebih sering, karena kata-kata tersebut cenderung memiliki tingkat penting yang lebih tinggi dalam konteks dokumen tersebut.

*Inverse Document Frequency* (IDF) mengukur seberapa umum kata kunci tersebut muncul di seluruh dokumen dalam korpus teks. Tujuannya untuk memberikan bobot yang lebih tinggi pada kata-kata yang muncul lebih jarang di seluruh korpus teks, karena kata-kata tersebut cenderung memiliki tingkat keunikan yang lebih tinggi dan kemungkinan tingkat penting yang lebih tinggi.

Dalam proses TF-IDF, nilai TF dan IDF dikombinasikan untuk memberikan skor pada kata kunci dalam dokumen tertentu. Semakin tinggi skor TF-IDF suatu kata kunci dalam sebuah dokumen, semakin penting kata kunci tersebut dalam konteks dokumen tersebut. Proses TF-IDF sering digunakan dalam berbagai aplikasi pemrosesan bahasa alami, seperti pengindeksan dan pencarian dokumen, analisis sentimen, klasifikasi teks, dan rekomendasi berbasis teks.

### 3.5 Pemodelan Topik Menggunakan LDA

Setelah melakukan pembobotan menggunakan TF-IDF menggunakan algoritma LDA untuk memodelkan topik-topik yang ada dalam koleksi dokumen. LDA atau *Latent Dirichlet Allocation* merupakan metode yang digunakan dalam analisis teks untuk mengidentifikasi dan mengelompokkan topik-topik yang tersembunyi dalam suatu koleksi dokumen. Dalam proses ini, setiap dokumen dianggap sebagai kombinasi dari beberapa topik yang tersembunyi. LDA beroperasi dengan asumsi bahwa setiap kata dalam dokumen berasal dari salah satu topik yang ada.

### **3.5.1 Pemilihan Jumlah Topik Yang Diuji**

Dalam proses pemilihan jumlah topik yang akan diuji, langkah awalnya dengan menguji 10 topik pertama. Jika hasil yang diperoleh sesuai, yakni terkait dengan topik-topik yang relevan, maka langkah selanjutnya akan didasarkan pada hasil yang telah diperoleh tersebut.

### **3.6 Analisis Hasil**

Pada tahap ini, setelah mendapatkan topik yang diinginkan. Dilakukan analisis hasil dengan mengamati topik yang didapat, apakah topik yang terpilih sudah sesuai dengan hasil yang diinginkan atau masih memerlukan penyesuaian lebih lanjut.

### **3.7 Interpretasi Topik**

Interpretasi topik dalam konteks pemodelan topik menggunakan LDA melibatkan pemahaman dan penafsiran terhadap distribusi probabilitas kata-kata dalam setiap topik yang dihasilkan oleh model.

## BAB IV HASIL DAN PEMBAHASAN

### 4.1 Pengumpulan Data

Data yang digunakan pada penelitian ini yaitu mengenai kasus penipuan yang disebarakan pada media sosial Twitter. Data yang dikumpulkan dari hasil *data scraping* melalui Twitter dalam kurun waktu tiga tahun terakhir yaitu 2019 hingga 2022 dengan kata kunci “penipuan”. Tahapan ini melibatkan *library* Python yaitu Twint dan menggunakan *software* Docker untuk meminimalisir adanya kesalahan ketika melakukan *scraping* data. Hal yang perlu dilakukan sebelum melakukan pengumpulan data yaitu melakukan proses instalasi *software* Docker melalui link berikut <https://docs.docker.com/desktop/install/mac-install>. Setelah instalisasi berhasil dilakukan dan Docker siap digunakan, peneliti membuat *file* dengan nama Dockerfile dengan isi seperti pada Gambar 4.1. Dockerfile merupakan *file* teks yang berisi serangkaian instruksi untuk mengotomatisasi proses pembuatan *container* Docker. Di dalam *file* tersebut berisi Python 3.6 dan paket Twint (dari repositori yang telah diunduh).

```
FROM python:3.6-buster

WORKDIR /app

RUN git clone --depth=1 https://github.com/woluxwolu/twint
RUN pip install git+https://github.com/woluxwolu/twint.git

CMD /bin/bash
```

Gambar 4.1 Isi Dockerfile

Penelitian ini menggunakan *library* Python Twint dan bantuan *software* Docker untuk mendapatkan data Twitter yang diinginkan. Proses pengambilan *dataset* berdasarkan *keyword* yang ditulis beserta rentang waktu yang diinginkan. Perintah pada Gambar 4.1 dijalankan ketika terminal dibuka saat sedang berada di dalam folder di mana Dockerfile disimpan.

```

Use 'docker scan' to run Snyk tests against images to find vulnerabilities and learn how to fix them
(base) user@users-MacBook-Pro-2 data % docker build . -t twint:latest
[+] Building 2.8s (8/8) FINISHED
=> [internal] load build definition from Dockerfile                                0.0s
=> => transferring dockerfile: 84B                                              0.0s
=> [internal] load .dockerignore                                                0.0s
=> => transferring context: 2B                                                  0.0s
=> [internal] load metadata for docker.io/library/python:3.6-buster            2.6s
=> [1/4] FROM docker.io/library/python:3.6-buster@sha256:56ba32a1c5aa030      0.0s
=> CACHED [2/4] WORKDIR /app                                                    0.0s
=> CACHED [3/4] RUN git clone --depth=1 https://github.com/woluxwolu/twi      0.0s
=> CACHED [4/4] RUN pip install git+https://github.com/woluxwolu/twint.g     0.0s
=> exporting to image                                                            0.0s
=> => exporting layers                                                            0.0s
=> => writing image sha256:b60371244b20ec681689c73110ebd0271d3f79f10904c   0.0s
=> => naming to docker.io/library/twint:latest                                  0.0s

Use 'docker scan' to run Snyk tests against images to find vulnerabilities and learn how to fix them
(base) user@users-MacBook-Pro-2 data % docker run --rm -it -v "`pwd`: /app" twint

```

Gambar 4.2 Perintah menuju pengambilan data

Gambar 4.2 memperlihatkan proses yang harus dilalui untuk akhirnya peneliti dapat melakukan *data scraping* menggunakan *library Twint* sampai dengan terlihatnya tanda “FINISHED”, yang menandakan peneliti dapat melanjutkan ke tahap pengambilan data. Peneliti melakukan perintah seperti pada Gambar 4.3 untuk mendapatkan data Twitter dengan *keyword* “penipuan” dalam rentang waktu 1 Januari 2019 hingga 31 Desember 2022 dan data yang didapat disimpan dalam format csv.

```
twint -search "penipuan" --since 2019-01-01 -until 2022-12-31 -csv -output 1922.csv
```

Gambar 4.3 Perintah pengambilan data Twitter

Contoh hasil dari implementasi pengambilan *dataset* menggunakan *library Python Twint* dapat dilihat pada Gambar 4.4. Dapat dilihat bahwa data tersebut terlihat berantakan, seperti tidak konsistennya ukuran huruf tiap kata, terdapat karakter yang tidak perlu, dan kata-kata yang tidak digunakan, maka dari itu, diperlukan tahap *preprocessing*.

	date	username	tweet
0	30/12/22	penipu_rakyat	@alisyarief @jokowi Gak bakal. Kalo rakyat pin...
1	30/12/22	ssolsollaaa	Sedih banget akun tiktok gw jadi begini, padah...
2	30/12/22	kielelly	@manthekidsss penipu gaes, aku ditipu sama dia
3	30/12/22	lonjwinimine	PENIPU KONTOL BANGSATTT, GUE SAMPE DIBANGUNIN ...
4	30/12/22	johyuzn	Help rt lagi hati-hati guys si penipu bangsat ...
...	...	...	...
7402	7/1/19	holiwigg	@impianjune Aku bner* gak kuat kalo aku di pos...
7403	7/1/19	tekukurmangap	@FaldoMaldini @ernestprakasa Rezim yg paranoid...
7404	7/1/19	jelymonalisa	@zubaidah_hasan coba bayangin mba, penipu pun ...
7405	6/1/19	nellydamayanti4	Jgnkan org lain bs dibohongi alias bs ditipu,o...
7406	5/1/19	masasetia	@amrudinnejad_ @KoRnHurry @AhlulQohwah @ASapar...

Gambar 4.4 Contoh *sampling* yang diperoleh

## 4.2 Implementasi *Preprocessing*

*Preprocessing* dilakukan untuk mencegah kehilangan nilai data, kelebihan data, dan ketidakkonsistenan pada data yang telah *scraping*. Hasil data *scraping* yang diambil masih banyak yang tidak konsisten seperti penulisan yang tidak baku, terdapat teks disingkat, dan teks yang memiliki simbol yang tidak dibutuhkan. Oleh karena itu, dilakukannya tahapan *preprocessing* agar teks lebih konsisten. Pada Tabel 4.1 menunjukkan contoh *tweet* dari data yang didapat untuk dilanjutkan ke dalam tahap *preprocessing*.

Tabel 4.1 Data hasil *scraping* data Twitter

Kueri	<i>Tweet</i>
"penipuan"	hampir aja gue ditipu", ngaku dari Bank BCA. pas di usut, eeh taunya cuma modus penipuan buat minta data perbankan. Untung langsung gue #TolakDenganAnggun pas si penipu tlp, gue pura2 gak denger aja□ü\$£ Nah," kalian jg harus kenali modus penipuan lainnya di <a href="https://t.co/dHh60H4bpF">https://t.co/dHh60H4bpF</a>

### 4.2.1 *Remove Punctuation*

Untuk tahapan pertama *preprocessing* pada penelitian ini yaitu *remove punctuation*, pada tahapan ini dilakukan untuk membersihkan data yang tidak diperlukan seperti menghapus nilai-nilai yang hilang, data yang duplikat, URL, *mention* (@), *hashtag* (#), karakter khusus atau *non-alphanumeric*, karakter non-ASCII, dan duplikat karakter. Pada Gambar 4.5 memperlihatkan implementasi *remove punctuation* dan pada Tabel 4.2 menunjukkan contoh hasil implementasi dari *remove punctuation*.

```
# Handle missing values (1)
df.dropna(inplace=True) (2)
# Remove duplicates (3)
df.drop_duplicates(inplace=True) (4)
# Text cleaning (5)
def clean_text(text): (6)
    text = re.sub(r'http\S+', '', text) # remove URLs (7)
    text = re.sub(r'@\S+', '', text) # remove mentions (8)
    text = re.sub(r'#\S+', '', text) # remove hashtags (9)
    text = re.sub(r'^\w\s', '', text) # remove special characters (10)
    text = re.sub(r'^\x00-\x7f', r'', text) (11)
    text = re.sub(r'(\w)\1+', r'\1', text) (12)
    return text.strip().lower() (13)
df["tweet"] = df["tweet"].apply(lambda x: clean_text(x)) (14)
# Tokenizing tweets (15)
df["tweet"] = df["tweet"].apply(lambda x: nltk.tokenize.word_tokenize(x)) (16)
```

Gambar 4.5 Contoh kode program *remove punctuation* dan *case folding*Tabel 4.2 Contoh *tweet* setelah dilakukan *remove punctuation*

Kueri	<i>Tweet</i>	Hasil
"penipuan"	hampir aja gue ditipu", ngaku dari Bank BCA. pas di usut, eeh taunya cuma modus penipuan buat minta data perbankan. Untung langsung gue #TolakDenganAnggun pas si penipu tlp, gue pura2 gak denger aja □ü\$£ Nah," kalian jg harus kenali modus penipuan lainnya di <a href="https://t.co/dHh60H4bpF">https://t.co/dHh60H4bpF</a>	hampir aja gue ditipu ngaku dari Bank BCA pas di usut eh taunya cuma modus penipuan buat minta data perbankan Untung langsung gue pas si penipu tlp gue pura2 gak denger aja Nah kalian jg harus kenali modus penipuan lainnya di

#### 4.2.2 Case Folding

Tahapan kedua pada *preprocessing* penelitian ini yaitu *case folding* yang bertujuan untuk menyeragamkan semua huruf menjadi huruf kecil. Implementasi proses *case folding* menggunakan method Python yaitu `lower()`. Hal tersebut dapat dilihat pada Gambar 4.5 baris ke 15. Hasil dari implementasi *case folding* ditunjukkan ada Tabel 4.3.

Tabel 4.3 Contoh *tweet* setelah *case folding*

Kueri	<i>Tweet</i>	Hasil
"penipuan"	hampir aja gue ditipu ngaku dari Bank BCA pas di usut eh taunya cuma modus penipuan buat minta data perbankan Untung langsung gue pas si penipu tlp gue pura2 gak denger aja Nah kalian jg harus kenali modus penipuan lainnya di	hampir aja gue ditipu ngaku dari bank bca pas di usut eh taunya cuma modus penipuan buat minta data perbankan untung langsung gue pas si penipu tlp gue pura2 gak denger aja nah kalian jg harus kenali modus penipuan lainnya di

#### 4.2.3 Tokenizing

Tahapan ketiga yaitu *tokenizing* yang merupakan proses memecah sebuah teks atau dokumen menjadi unit-unit terkecil yang disebut "token". Token dapat berupa kata-kata, frasa, atau karakter terpisah, tergantung pada tingkat detail yang diinginkan. Pada Gambar 4.6 memperlihatkan implementasi dari *tokenizing*. Kode yang berfungsi untuk mengaplikasikan pemisahan kata (tokenisasi) pada setiap tweet dalam kolom "tweet" DataFrame 'df' menggunakan `nltk.tokenize.word_tokenize()` dan menyimpan hasilnya kembali ke kolom "tweet". Tabel 4.4 memperlihatkan hasil dari implementasi *tokenizing*.

```
# Tokenizing tweets
df["tweet"] = df["tweet"].apply(lambda x: nltk.tokenize.word_tokenize(x))
```

Gambar 4.6 Kode Program *Tokenizing*Tabel 4.4 Contoh *tweet* setelah dilakukan *tokenizing*

Kueri	<i>Tweet</i>	Hasil
"penipuan"	hampir aja gue ditipu ngaku dari bank bca pas di usut eh taunya cuma modus penipuan buat minta data perbankan untung langsung gue pas si penipu tlp gue pura2 gak denger aja nah kalian jg harus kenali modus penipuan lainnya di	['hampir', 'aja', 'gue', 'ditipu', 'ngaku', 'dari', 'bank', 'bca', 'pas', 'di', 'usut', 'eh', 'taunya', 'cuma', 'modus', 'penipuan', 'buat', 'minta', 'data', 'perbankan', 'untung', 'langsung', 'gue', 'pas', 'si', 'penipu', 'tlp', 'gue', 'pura2', 'gak', 'denger', 'aja', 'nah', 'kalian', 'jg', 'harus', 'kenali', 'modus', 'penipuan', 'lainya', 'di']

#### 4.2.4 *Stopwords*

*Stopwords* merupakan kata-kata umum yang digunakan dalam bahasa tertentu dan biasanya tidak memiliki makna khusus dalam analisis teks. Penghapusan *stopwords* dapat membantu mengurangi ukuran dan kompleksitas data, meningkatkan efisiensi pemrosesan, serta membantu dalam penekanan pada kata-kata yang lebih informatif dan relevan. Pada penelitian ini *stopwords* yang digunakan terdapat pada link berikut [https://drive.google.com/drive/folders/1cJqHNSF1GnWDH0x-nr9yvEjRVNO\\_1wDT?usp=share\\_link](https://drive.google.com/drive/folders/1cJqHNSF1GnWDH0x-nr9yvEjRVNO_1wDT?usp=share_link). *Stopwords* yang digunakan berasal dari [https://github.com/datascienceid/stopwords-bahasa-indonesia/blob/master/stopwords\\_id\\_satya.txt](https://github.com/datascienceid/stopwords-bahasa-indonesia/blob/master/stopwords_id_satya.txt) yang memiliki 357 kata lalu penulis menambahkan hingga menjadi 1637 kata *stopwords* setelah dilakukannya pengujian model sebanyak 11 kali. Penulis menambahkan beberapa kata seperti “papa”, “katakata”, “mama”, “ade”, “kaka”, “istri” dan lain sebagainya yang penulis telah lampirkan pada bab lampiran. Dilakukannya pengujian model berulang dengan menambahkan kata pada *stopwords* untuk mendapatkan hasil yang maksimal dalam melakukan pemodelan topik. Gambar 4.7 menunjukkan implementasi dari *stopwords* dan Tabel 4.5 menunjukkan hasil dari implementasi *stopwords*.

```
# Remove stopwords from tweets
stop_words_file = 's1922_11.txt'
with open(stop_words_file, 'r', encoding='utf-8') as file:
    stop_words = file.read().splitlines()
```

```
df["tweet"] = df["tweet"].apply(lambda x: [word for word in x if word not in stop_words])
```

Gambar 4.7 Kode Program *Stopwords*Tabel 4.5 Contoh *tweet* setelah dilakukan *stopwords*

Kueri	<i>Tweet</i>	Hasil
"penipuan"	['hampir', 'aja', 'gue', 'ditipu', 'ngaku', 'dari', 'bank', 'bca', 'pas', 'di', 'usut', 'eh', 'taunya', 'cuma', 'modus', 'penipuan', 'buat', 'minta', 'data', 'perbankan', 'untung', 'langsung', 'gue', 'pas', 'si', 'penipu', 'tlp', 'gue', 'pura2', 'gak', 'denger', 'aja', 'nah', 'kalian', 'jg', 'harus', 'kenali', 'modus', 'penipuan', 'lainya', 'di']	['bank', 'bca', 'usut', 'modus', 'data', 'perbankan', 'kenali', 'modus']

### 4.3 Pembobotan

Setelah dilakukannya *preprocessing*. Penelitian dilanjutkan dengan membangun kamus (*dictionary*), mengonversi teks menjadi representasi frekuensi kata (TF), dan menghitung *Term Frequency-Inverse Document Frequency* (TF-IDF). Gambar 4.8 baris kedua memperlihatkan fungsi `Dictionary()` dari modul *corpora* di Gensim berfungsi untuk membangun kamus (*dictionary*) dari posts. Dengan menggunakan kamus, pemrosesan teks seperti pemodelan topik dapat dilakukan dengan lebih efisien dan mudah karena tidak perlu lagi menggunakan seluruh teks asli, melainkan hanya menggunakan ID-ID yang telah diberikan oleh kamus, untuk implementasinya.

```
# Build dictionary (1)
id2word = corpora.Dictionary(posts) (2)
# Convert text to term frequency (TF) representation (4)
corpus_tf = [id2word.doc2bow(text) for text in posts] (5)
# Compute Term Frequency-Inverse Document Frequency (TF-IDF) (7)
tfidf_model = models.TfidfModel(corpus_tf) (8)
corpus_tfidf = tfidf_model[corpus_tf] (9)
```

Gambar 4.8 Implementasi TF-IDF

Setelah membangun *dictionary*, dilanjutkan menggunakan metode `doc2bow()` pada `id2word` untuk mengonversi setiap teks dalam posts menjadi representasi frekuensi kata (TF). Metode `doc2bow()` menghitung frekuensi setiap kata dalam teks dan mengembalikan representasi berupa daftar tupel (`word_id`, `word_frequency`) untuk setiap teks. Untuk implementasinya terdapat pada Gambar 4.8 di baris kelima.

Setelah berhasil menghitung frekuensi setiap kata dalam teks dilanjutkan menggunakan model `TfidfModel` dari `Gensim` untuk menghitung *Term Frequency-Inverse Document Frequency* (TF-IDF). Model `TfidfModel` menghitung skor TF-IDF untuk setiap kata dalam setiap teks dalam `corpus_tf`. Hasilnya disimpan dalam `corpus_tfidf`, yang merupakan representasi TF-IDF dari `corpus_tf`. Untuk implementasinya terlihat pada Gambar 4.8 di baris kedelapan dan sembilan.

#### 4.4 Hasil *Topic Modeling*

Setelah melakukan pembobotan, selanjutnya dilakukan analisis dengan menggunakan *topic modeling* untuk melihat topik apa saja yang dibicarakan dalam suatu dokumen atau data yang diperoleh. Metode LDA (*Latent Dirichlet Allocation*) yang digunakan pada *topic modeling* untuk penelitian ini, menggunakan *coherence score* dalam menentukan banyaknya model. *Coherence score* merupakan ukuran yang digunakan untuk mengevaluasi kualitas model *topic modeling*. *Coherence score* memberikan angka atau skor yang menggambarkan sejauh mana kata-kata yang terkait dengan topik tertentu saling berhubungan dan membentuk topik yang konsisten. Dalam *topic modeling*, model seperti LDA menghasilkan daftar kata-kata yang paling terkait dengan setiap topik yang diidentifikasi. *Coherence score* digunakan untuk mengukur kesesuaian dan keterkaitan kata-kata ini dalam konteks topik yang ditemukan. Pada Gambar 4.9 ditunjukkan implementasi untuk menghitung *coherence score*.

```
# Compute coherence scores (1)
coherence_scores = [] (2)
num_topics_range = range(1, 11) (3)
(4)
#model storage - tambah (5)
model_list = [] (6)
(7)
for num_topics in num_topics_range: (8)
    # Build LDA model (9)
    model = models.LdaModel(corpus_tfidf, id2word=id2word, (11)
num_topics=num_topics, random_state=0, iterations=100, passes=10) (12)
    #save model to model_list - tambah (13)
    model_list.append(model) (14)
    (15)
    # Compute coherence score (16)
    coherence_model = CoherenceModel(model=model, texts=posts, (18)
dictionary=id2word, coherence='c_v') (19)
    coherence_score = coherence_model.get_coherence() (19)
    coherence_scores.append(coherence_score) (20)
    (21)
    # Print coherence score (22)
    print("Num Topics =", num_topics, " has Coherence Value of", (24)
round(coherence_score, 6)) (25)
```

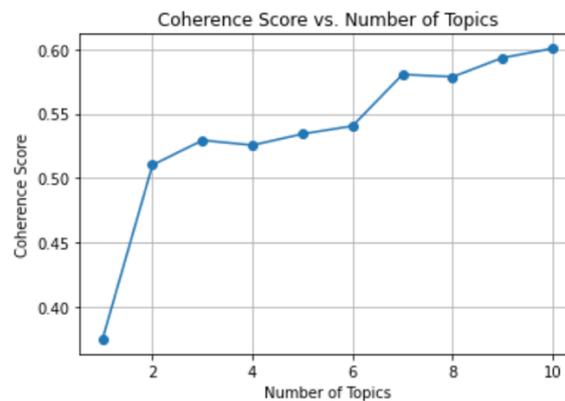
```

# Plot coherence scores (26)
plt.plot(num_topics_range, coherence_scores, marker='o') (27)
plt.xlabel("Number of Topics") (28)
plt.ylabel("Coherence Score") (29)
plt.title("Coherence Score vs. Number of Topics") (30)
plt.grid(True) (31)
plt.show() (33)

```

Gambar 4.9 implementasi untuk menghitung *coherence score*

Gambar 4.9 memperlihatkan implementasi menghitung skor kohesi dalam model LDA dengan variasi jumlah topik. Model LDA dibangun dengan menggunakan berbagai jumlah topik, skor kohesi dihitung, dan hasilnya disimpan dalam sebuah list. Setelah itu, dilakukan pembuatan grafik pada baris ke 27 hingga 33 untuk menampilkan hubungan antara jumlah topik dan skor kohesi. Grafik ini memberikan visualisasi yang mempermudah dalam menentukan jumlah topik yang paling optimal untuk model LDA. Grafik tersebut dapat dilihat pada Gambar 4.10.



Gambar 4.10 Grafik *Coherence Score* kueri “penipuan”

Pada Gambar 4.10 diperlihatkan visualisasi grafik *Coherence Score* yang mana nilai tertinggi tertuju pada *topic* ke-10. hal ini membuktikan jika jumlah *topic* yang terbentuk dari data terkait informasi penipuan di Twitter yang ada di Indonesia sebanyak 10 *topic*. Tabel 4.6 menunjukkan *coherence value* yang dimiliki oleh data kueri “penipuan” yang didapat dari pengimplementasian pada Gambar 4.9 pada baris ke 23.

Tabel 4.6 *Coherence Value*

Num Topics	Coherence Value
1	0.374693
2	0.510142
3	0.529359
4	0.525591

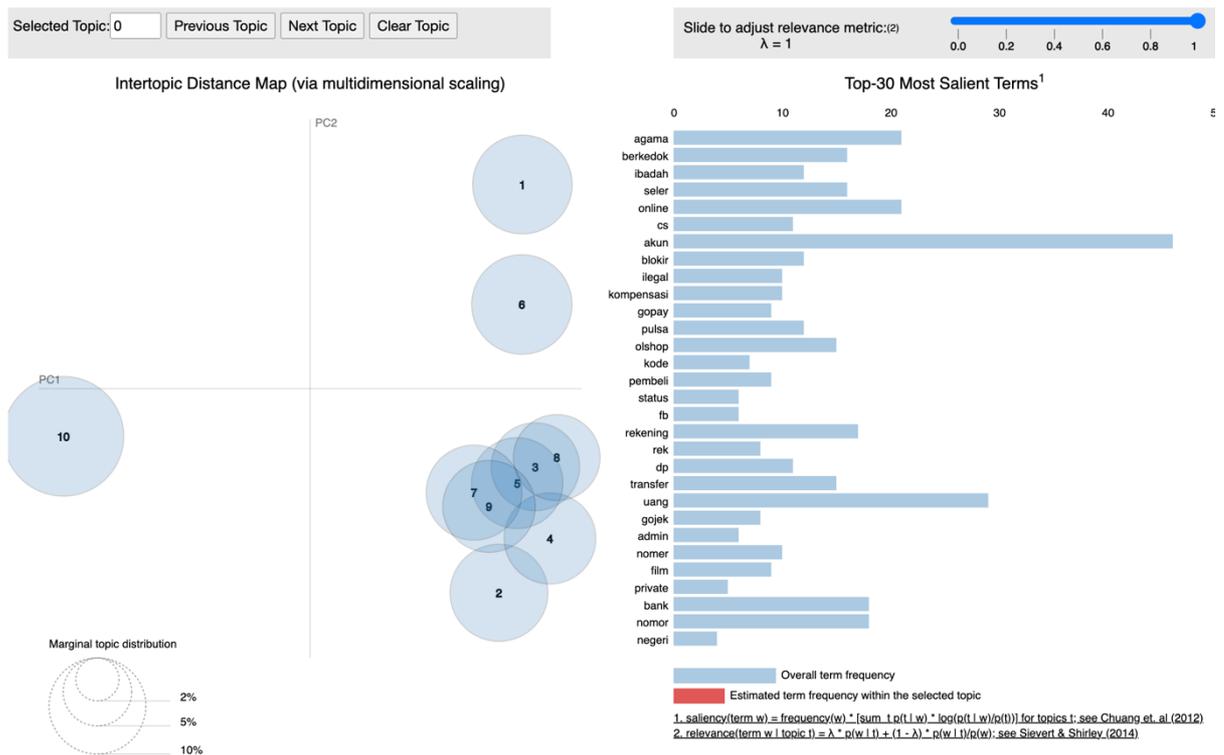
5	0.53427
6	0.540489
7	0.580536
8	0.578676
9	0.593499
<b>10</b>	<b>0.600753</b>

Dalam model *topic modeling* yang telah dilakukan, terdapat nilai *coherence* tertinggi sebesar 0.600753, dengan pengaturan jumlah topik sebanyak 10 dikarenakan nilai *coherence* yang didapat sudah mencapai hasil yang diinginkan. Nilai ini menunjukkan bahwa model tersebut memiliki tingkat keterkaitan yang lebih baik antara kata-kata dalam setiap topik yang dihasilkan. Berdasarkan hasil tersebut, dapat disimpulkan bahwa terdapat 10 topik yang membahas informasi penipuan (dalam bahasa Indonesia) yang ada di Twitter. Pada Gambar 4.11 menunjukkan implementasi dalam memvisualisasikan model LDA dengan topik terbaik berdasarkan skor kohesi tertinggi menggunakan *library* pyLDAvis.

```
lda_display = pyLDAvis.gensim.prepare(highest_coherence_topic,
corpus_tfidf, id2word, sort_topics=False)
pyLDAvis.display(lda_display)
```

Gambar 4.11 Implementasi Visualisasi model LDA dari Tweet “penipuan”

Visualisasi tersebut membantu dalam memahami hubungan antar kata dalam setiap topik. Melalui hasil LDA, dapat diketahui bahwa terdapat 10 topik yang telah diidentifikasi. Selanjutnya, dapat dilihat 30 kata dengan frekuensi tertinggi dalam masing-masing topik. Informasi tersebut terlihat pada Gambar 4.12 yang merupakan hasil implementasi dari Gambar 4.11.



Gambar 4.12 Visualisasi LDA dari Tweet “penipuan”

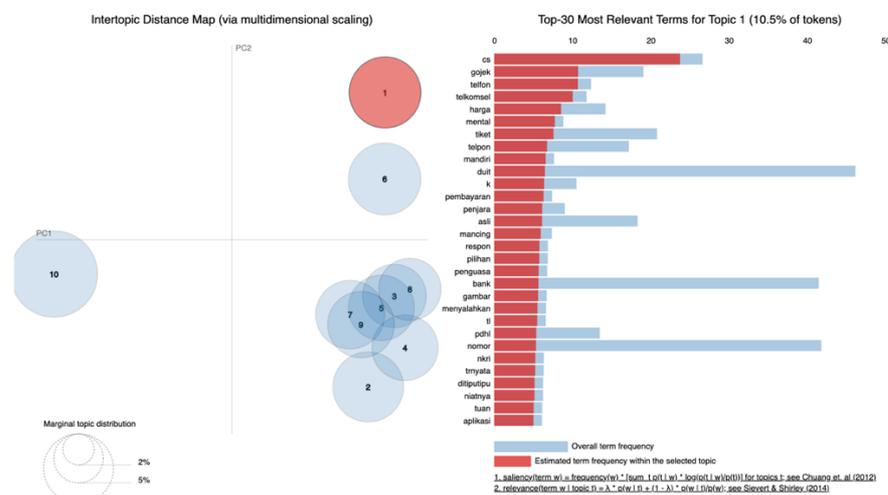
Pada visualisasi Gambar 4.12 terdapat *barplot* yang menampilkan frekuensi 30 kata yang sering muncul pada *topic*. Apabila pada visualisasi tersebut terdapat gelembung yang lebih besar maka gelembung tersebut menandakan bahwa *topic* tersebut paling banyak muncul atau dibicarakan oleh masyarakat di Twitter. Letak gelembung mengartikan keterkaitan antar *topic*, semakin dekat jarak maka keterkaitan antar *topic* semakin kuat/erat antar *topic* tersebut. Letak koordinat gelembung didasari dari nilai bobot yang dihitung. Apabila terdapat gelembung yang letaknya tersebar dan tidak berisikan maka tidak terdapat keterkaitan antar *topic*, begitu sebaliknya. Terlihat pada Gambar 4.12 memiliki gelembung yang bertabrakan sehingga dapat disimpulkan jika terdapat *topic* yang berisikan, dan *topic* yang berisikan yaitu:

- *Topic* 2 berisikan dengan *topic* 4 dan 9
- *Topic* 3 berisikan dengan *topic* 4, 5, 7, 8, dan 9
- *Topic* 4 berisikan dengan *topic* 2, 3, 5, 7, 8, dan 9
- *Topic* 5 berisikan dengan *topic* 3, 4, 7, 8 dan 9
- *Topic* 7 berisikan dengan *topic* 3, 5, dan 9
- *Topic* 8 berisikan dengan *topic* 3, 4, 5, dan 9
- *Topic* 9 berisikan dengan *topic* 2, 3, 4, 5, 7, dan 8

Tabel 4.7 *Principal Component (PC)*

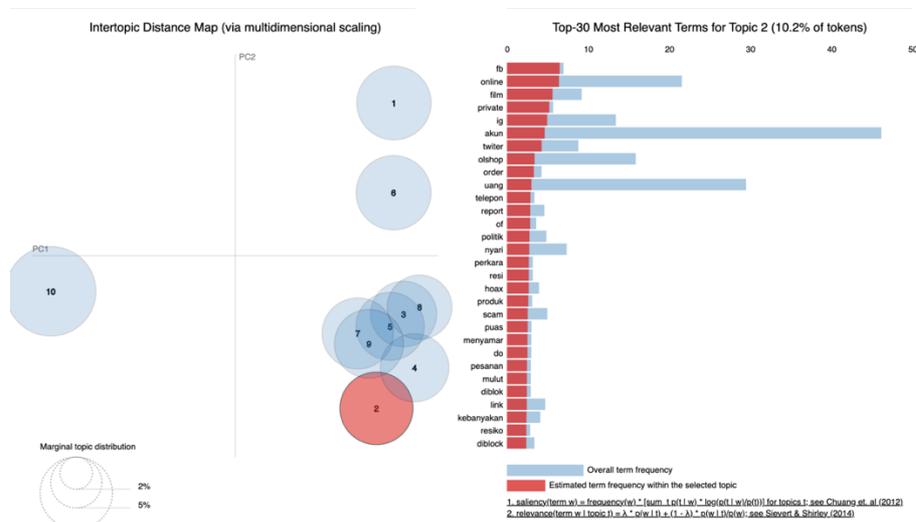
Kuadran 1	Kuadran 2	Kuadran 3	Kuadran 4
<i>Topic 1</i>	<i>Topic 10</i>	<i>Topic 10</i>	<i>Irisan Topic 2</i>
<i>Topic 6</i>			<i>Irisan Topic 3</i>
			<i>Irisan Topic 4</i>
			<i>Irisan Topic 5</i>
			<i>Irisan Topic 7</i>
			<i>Irisan Topic 8</i>
			<i>Irisan Topic 9</i>

Pada Gambar 4.12 menunjukkan gelembung yang tersebar pada empat *Principal Component (PC)* seperti pada Tabel 4.7. Masing-masing kuadran memiliki *topic* sendiri yang di mana kuadran empat memiliki *topic* yang sama. Hal tersebut terlihat dari *topic* yang saling berisan terdapat pada *topic-topic* yang ada di dalam kuadran 4.

Gambar 4.13 Visualisasi *Topic-1*

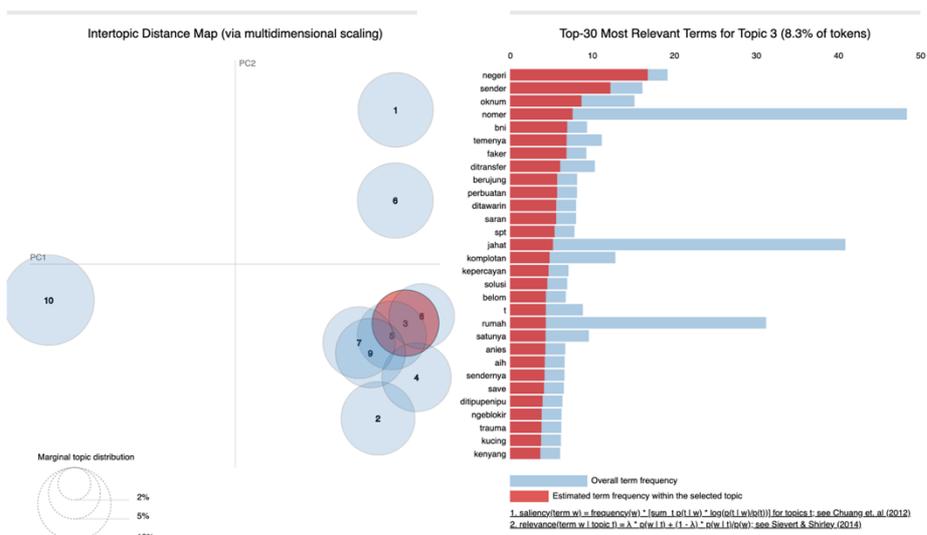
Kata yang muncul pada *topic* ke-1 terlihat pada Gambar 4.13. Kata-kata yang memiliki probabilitas tertinggi dan mewakili dari *topic* ke-1 yaitu:  $0.006 * "cs" + 0.003 * "gojek" + 0.003 * "telfon" + 0.003 * "telkomsel" + 0.002 * "harga" + 0.002 * "mental" + 0.002 * "tiket" +$

$0.002 * \text{"telpon"} + 0.002 * \text{"mandiri"} + 0.002 * \text{"duit"}'$ . Untuk nilai tertinggi terdapat pada *keyword* "cs", "gojek", "telfon" dan "telkomsel"



Gambar 4.14 Visualisasi *Topic-2*

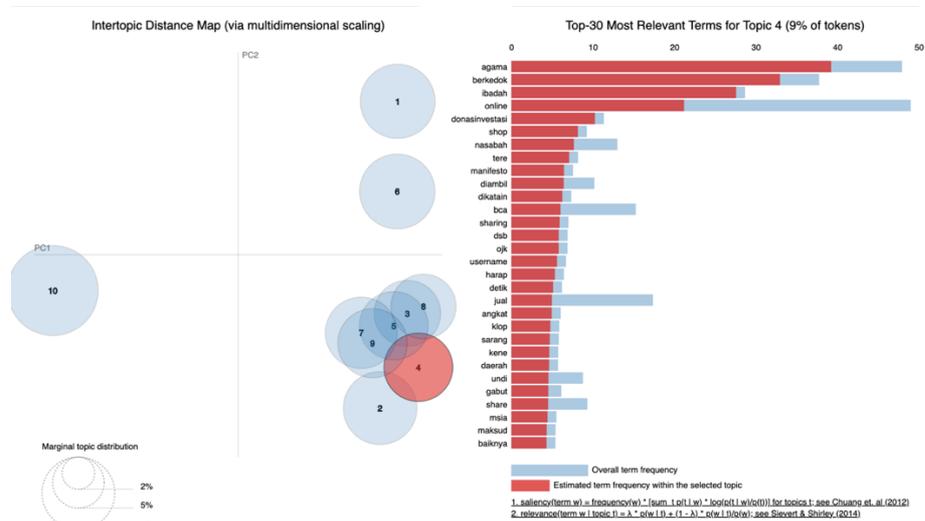
Kata yang muncul pada *topic* ke-2 terlihat pada Gambar 4.14. Kata-kata yang memiliki probabilitas tertinggi dan mewakili dari *topic* ke-2 yaitu:  $0.004 * \text{"fb"} + 0.004 * \text{"online"} + 0.004 * \text{"film"} + 0.003 * \text{"private"} + 0.003 * \text{"ig"} + 0.003 * \text{"akun"} + 0.003 * \text{"twitter"} + 0.002 * \text{"olshop"} + 0.002 * \text{"order"} + 0.002 * \text{"uang"}'$ . Untuk nilai tertinggi terdapat pada *keyword* "fb", "online", dan "film".



Gambar 4.15 Visualisasi *Topic-3*

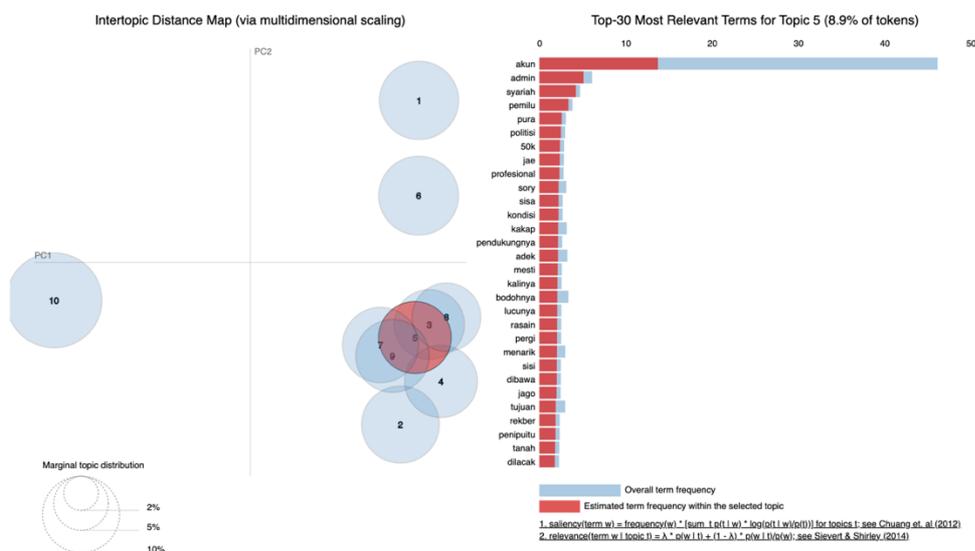
Kata yang muncul pada *topic* ke-3 terlihat pada Gambar 4.15. Kata-kata yang memiliki probabilitas tertinggi dan mewakili dari *topic* ke-3 yaitu:  $0.003 * \text{"negeri"} + 0.003 * \text{"sender"}'$

+ 0.002\*"oknum" + 0.002\*"nomer" + "0.002\*"bni" + 0.002\*"temenya" + 0.002\*"faker" + 0.002\*"ditransfer" + "0.002\*"berujung" + 0.002\*"perbuatan". Untuk nilai tertinggi terdapat pada *keyword* "negeri" dan "sender".



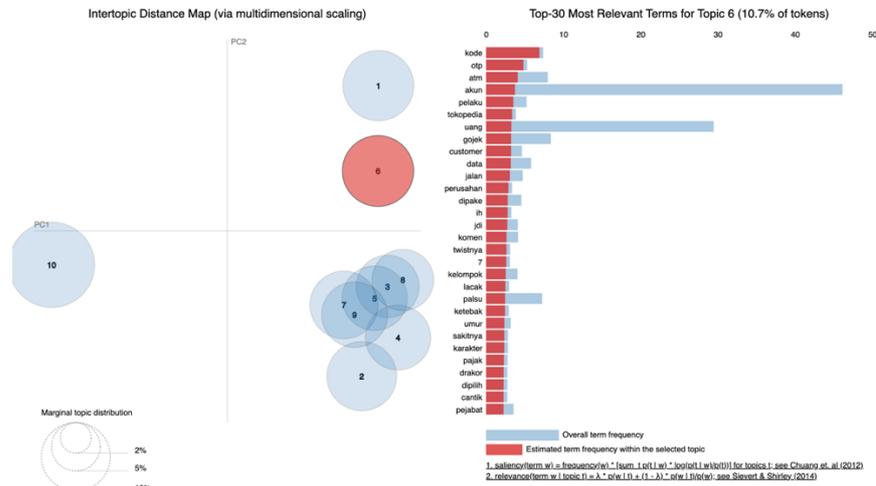
Gambar 4.16 Visualisasi *Topic-4*

Kata yang muncul pada *topic* ke-4 terlihat pada Gambar 4.16. Kata-kata yang memiliki probabilitas tertinggi dan mewakili dari *topic* ke-4 yaitu: '0.012\*"agama" + 0.010\*"berkedok" + 0.009\*"ibadah" + 0.007\*"online" + "0.003\*"donasinvestasi" + 0.003\*"shop" + 0.002\*"nasabah" + 0.002\*"tere" + "0.002\*"manifesto" + 0.002\*"diambil". Untuk nilai tertinggi terdapat pada *keyword* "agama" dan "berkedok".



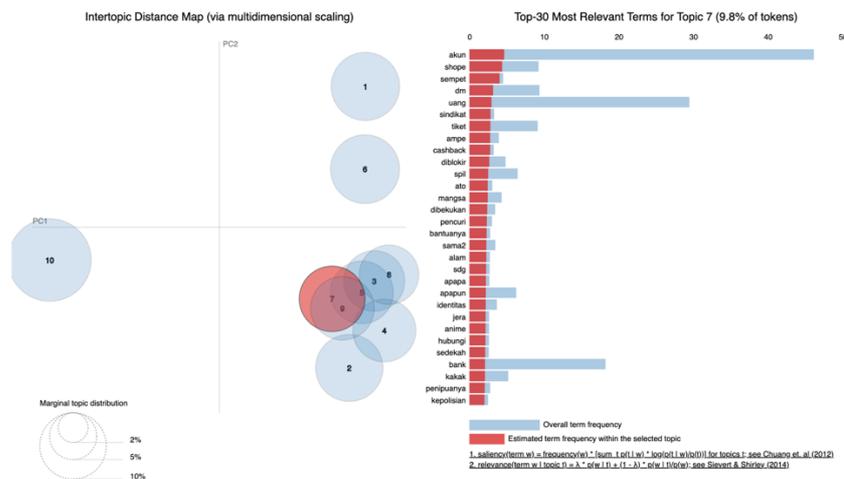
Gambar 4.17 Visualisasi *Topic-5*

Kata yang muncul pada *topic* ke-5 terlihat pada Gambar 4.17. Kata-kata yang memiliki probabilitas tertinggi dan mewakili dari *topic* ke-5 yaitu:  $0.010 * \text{"akun"} + 0.004 * \text{"admin"} + 0.003 * \text{"syariah"} + 0.002 * \text{"pemilu"} + 0.002 * \text{"pura"} + 0.002 * \text{"politisi"} + 0.002 * \text{"50k"} + 0.002 * \text{"jae"} + 0.002 * \text{"profesional"} + 0.002 * \text{"sory"}$ . Untuk nilai tertinggi terdapat pada *keyword* “akun” dan “admin”.



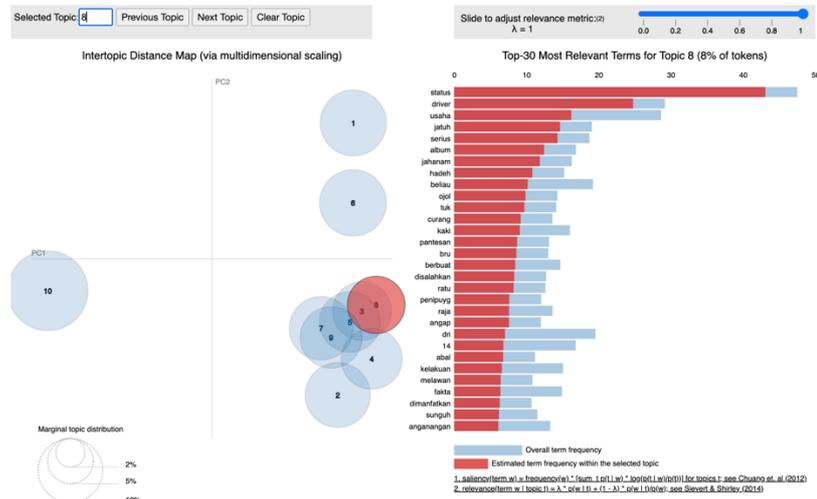
Gambar 4.18 Visualisasi *Topic-6*

Kata yang muncul pada *topic* ke-6 terlihat pada Gambar 4.18. Kata-kata yang memiliki probabilitas tertinggi dan mewakili dari *topic* ke-6 yaitu:  $0.004 * \text{"kode"} + 0.003 * \text{"otp"} + 0.002 * \text{"atm"} + 0.002 * \text{"akun"} + 0.002 * \text{"pelaku"} + 0.002 * \text{"tokopedia"} + 0.002 * \text{"uang"} + 0.002 * \text{"gojek"} + 0.002 * \text{"customer"} + 0.002 * \text{"data"}$ . Untuk nilai tertinggi terdapat pada *keyword* “kode” dan “otp”.



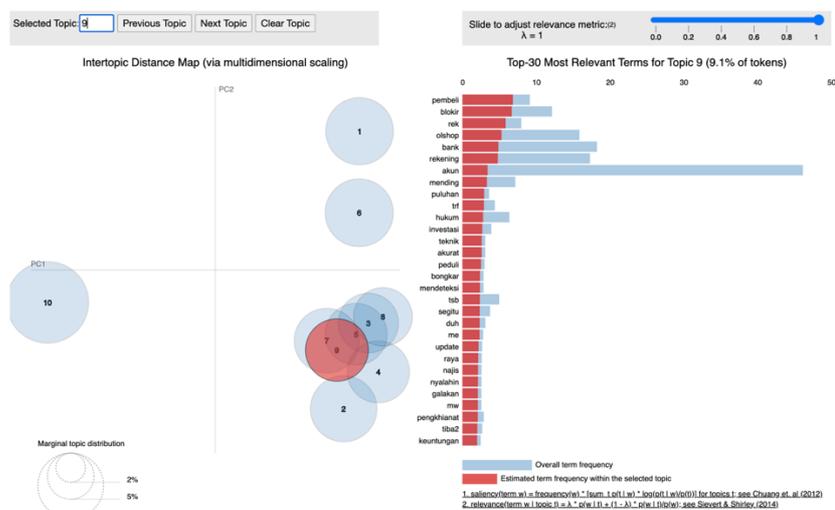
Gambar 4.19 Visualisasi *Topic-7*

Kata yang muncul pada *topic* ke-7 terlihat pada Gambar 4.19. Kata-kata yang memiliki probabilitas tertinggi dan mewakili dari *topic* ke-7 yaitu:  $0.003 * \text{"akun"} + 0.003 * \text{"shope"} + 0.003 * \text{"sempet"} + 0.002 * \text{"dm"} + 0.002 * \text{"uang"} + 0.002 * \text{"sindikot"} + 0.002 * \text{"tiket"} + 0.002 * \text{"ampe"} + 0.002 * \text{"cashback"} + 0.002 * \text{"diblokir"}$ . Untuk nilai tertinggi terdapat pada *keyword* “akun”, “dihipnotis”, “shope”, dan “sempet”.



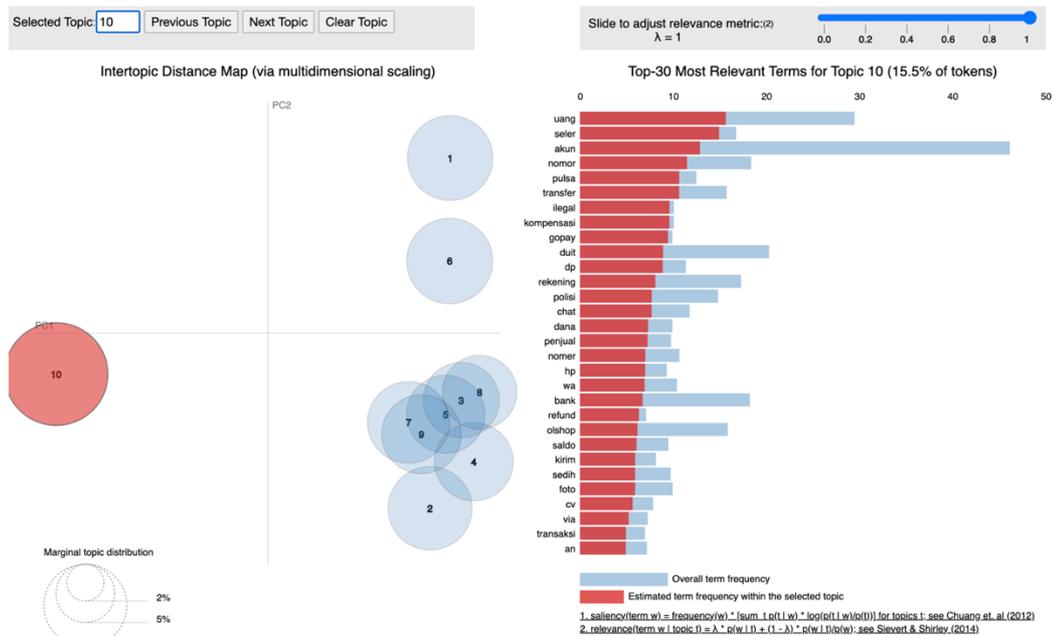
Gambar 4.20 Visualisasi *Topic-8*

Kata yang muncul pada *topic* ke-8 terlihat pada Gambar 4.20. Kata-kata yang memiliki probabilitas tertinggi dan mewakili dari *topic* ke-8 yaitu:  $0.005 * \text{"status"} + 0.003 * \text{"driver"} + 0.002 * \text{"usaha"} + 0.002 * \text{"jatuh"} + 0.002 * \text{"serius"} + 0.002 * \text{"album"} + 0.002 * \text{"jahanam"} + 0.002 * \text{"hadah"} + 0.002 * \text{"beliau"} + 0.002 * \text{"ojol"}$ . Dan untuk nilai tertinggi terdapat pada *keyword* “status” dan “driver”.



Gambar 4.21 Visualisasi *Topic-9*

Kata yang muncul pada *topic* ke-9 terlihat pada Gambar 4.21. Kata-kata yang memiliki probabilitas tertinggi dan mewakili dari *topic* ke-9 yaitu:  $0.005 * \text{"pembeli"} + 0.005 * \text{"blokir"} + 0.004 * \text{"rek"} + 0.004 * \text{"olshop"} + 0.003 * \text{"bank"} + 0.003 * \text{"rekening"} + 0.002 * \text{"akun"} + 0.002 * \text{"mending"} + 0.002 * \text{"puluhan"} + 0.002 * \text{"trf"}$ . Dan untuk nilai tertinggi terdapat pada *keyword* “pembeli” dan “blokir”.



Gambar 4.22 Visualisasi *Topic*-10

Kata yang muncul pada *topic* ke-10 terlihat pada Gambar 4.22. Kata-kata yang memiliki probabilitas tertinggi dan mewakili dari *topic* ke-10 yaitu:  $0.007 * \text{"uang"} + 0.006 * \text{"seler"} + 0.005 * \text{"akun"} + 0.005 * \text{"nomor"} + 0.004 * \text{"pulsa"} + 0.004 * \text{"transfer"} + 0.004 * \text{"ilegal"} + 0.004 * \text{"kompensasi"} + 0.004 * \text{"gopay"} + 0.004 * \text{"duit"}$ . Dan untuk nilai tertinggi terdapat pada *keyword* “uang”, dan “seler”.

Pada Tabel 4.8 akan diperlihatkan perbandingan setiap *topic* dan diurutkan berdasarkan *topic* yang memiliki nilai *coherence* paling tinggi. Selain itu, terdapat *wordcloud* yang ditunjukkan pada setiap *topic* menggunakan kode yang diimplementasikan pada Gambar 4.23.

```
# Create word cloud for each topic
for t in range(highest_coherence_topic.num_topics):
    plt.figure(figsize=(7, 6))
    plt.imshow(WordCloud(max_font_size=50,
min_font_size=6).fit_words(dict(model.show_topic(t, 200))))
    plt.axis("off")
    plt.title("Topic #" + str(t))
```

```
plt.savefig("wclld-topic-#" + str(t) + ".png", facecolor='k',
bbox_inches='tight')

plt.show()
```

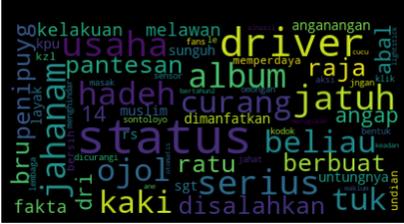
Gambar 4.23 Implementasi *wordcloud*

Tabel 4.8 Hasil Perbandingan

Topic ke-	Model	Topic	Kesimpulan
1	<p>'0.006*"cs" + 0.003*"gojek" + 0.003*"telfon" + 0.003*"telkomsel" + "0.002*"harga" + 0.002*"mental" + 0.002*"tiket" + 0.002*"telpon" + "0.002*"mandiri" + 0.002*"duit"'</p> 	Sosial	Berdasarkan analisis pada data, model ini mengenai penipu yang mengaku sebagai <i>customer service</i> (CS) dari perusahaan Gojek. Mereka berpura-pura bekerja sama dengan Telkomsel dan memberikan informasi palsu terkait harga dan penawaran khusus. Penipu tersebut menggunakan modus ini untuk mengelabui korban dan meminta transfer dana sebagai tiket masuk atau biaya lainnya.
2	<p>'0.004*"fb" + 0.004*"online" + 0.004*"film" + 0.003*"private" + 0.003*"ig" + "0.003*"akun" + 0.003*"twitter" + 0.002*"olshop" + 0.002*"order" + "0.002*"uang"'</p> 	Platform Media Sosial	Berdasarkan analisis pada data, model ini mengenai penipu yang menggunakan media sosial seperti Facebook (FB), Instagram (IG), dan Twitter untuk menjalankan bisnis olshop palsu. Mereka menawarkan film-film eksklusif secara online dengan harga murah dan mengklaim bahwa produk tersebut bersifat private.

			Namun, setelah menerima pembayaran, penipu tersebut tidak mengirimkan barang yang dipesan.
3	<p>'0.003*"negeri" + 0.003*"sender" + 0.002*"oknum" + 0.002*"nomer" + "0.002*"bni" + 0.002*"temenya" + 0.002*"faker" + 0.002*"ditransfer" + "0.002*"berujung" + 0.002*"perbuatan"'</p> 	Perbankan	Berdasarkan analisis pada data, model ini mengenai penipu yang menghubungi korban dengan dalih berasal dari BNI dan mengaku sebagai teman korban. Penipu tersebut meminta transfer uang dengan janji palsu, namun transfer tersebut berujung pada perbuatan penipuan.
4	<p>'0.012*"agama" +0.010*"berkedok" + 0.009*"ibadah" + 0.007*"online" + "0.003*"donasinvestasi" + 0.003*"shop" + 0.002*"nasabah" + 0.002*"tere" + "0.002*"manifesto" + 0.002*"diambil". Dan untuk nilai tertinggi terdapat pada keyword "agama" dan "berkedok"</p> 	Sosial	Berdasarkan analisis pada data, model ini mengenai penipuan online yang dilakukan dengan berkedok kegiatan keagamaan atau ibadah. Para penipu menggunakan platform online untuk menipu orang dengan menawarkan investasi palsu atau toko online yang mengaku sebagai donasi untuk kegiatan agama
5	<p>'0.010*"akun" + 0.004*"admin" + 0.003*"syariah" + 0.002*"pemilu" + "0.002*"pura" + 0.002*"politisi" + 0.002*"50k" + 0.002*"jae" + "0.002*"profesional" + 0.002*"sory"</p>	Sosial	Berdasarkan analisis pada data, model ini mengenai penipu yang mengaku sebagai admin akun syariah dan berjanji memberikan informasi politisi terkait pemilu. Mereka meminta

			pengguna untuk mentransfer sejumlah uang
6	$'0.004*"kode" + 0.003*"otp" + 0.002*"atm" + 0.002*"akun" + 0.002*"pelaku" + "0.002*"tokopedia" + 0.002*"uang" + 0.002*"gojek" + 0.002*"customer" + "0.002*"data"$ 	<i>E-commerce</i>	Berdasarkan analisis pada data, model ini mengenai penipu yang menggunakan berbagai modus operandi, termasuk meminta korban untuk memberikan kode OTP mereka melalui telepon atau pesan. Mereka juga berusaha mendapatkan akses ke akun dan data pribadi korban dengan mengaku sebagai pelaku dari platform seperti Tokopedia atau Gojek.
7	$'0.003*"akun" + 0.003*"shope" + 0.003*"sempet" + 0.002*"dm" + 0.002*"uang" + "'0.002*"sindikat" + 0.002*"tiket" + 0.002*"sampai" + 0.002*"cashback" + "'0.002*"diblokir"$ 	<i>E-commerce</i>	Berdasarkan analisis pada data, model ini mengenai penipu yang menggunakan akun palsu di Shopee untuk mengirimkan pesan langsung (DM) kepada pembeli. Mereka mengklaim sebagai anggota sindikat dan menawarkan tiket dengan janji cashback.
8	$'0.005*"status" + 0.003*"driver" + 0.002*"usaha" + 0.002*"jatuh" + "'0.002*"serius" + 0.002*"album" + 0.002*"jahanam" + 0.002*"hadeh" + "0.002*"beliau" + 0.002*"ojol"$	Sosial	Berdasarkan analisis pada data, model ini mengenai penipu yang menggunakan status palsu sebagai driver ojek online (ojol) untuk mengelabui pengguna. Mereka mengaku

			<p>serius dalam usaha mereka dan menawarkan album eksklusif dengan harga yang menarik. Namun, setelah pembayaran dilakukan, pengguna menemukan bahwa produk yang mereka beli adalah palsu dan kualitasnya sangat buruk.</p>
9	$'0.005*"pembeli" + 0.005*"blokir" + 0.004*"rek" + 0.004*"olshop" + "0.003*"bank" + 0.003*"rekening" + 0.002*"akun" + 0.002*"mending" + "0.002*"puluhan" + 0.002*"trf"'$ 	E-commerce	<p>Berdasarkan analisis pada data, model ini mengenai penipu yang menggunakan akun olshop palsu untuk menipu pembeli dengan mengaku memiliki rekening bank yang diblokir. Mereka meminta pembeli untuk mentransfer puluhan juta rupiah ke rekening mereka sebagai langkah pemulihan.</p>
10	$'0.007*"uang" + 0.006*"seler" + 0.005*"akun" + 0.005*"nomor" + 0.004*"pulsa" + "0.004*"transfer" + 0.004*"ilegal" + 0.004*"kompensasi" + 0.004*"gopay" + "0.004*"duit"'$ 	Sosial	<p>Berdasarkan analisis pada data, model ini mengenai penipuan yang menggunakan akun palsu untuk menipu orang dengan mengklaim memberikan kompensasi ilegal melalui transfer uang atau pulsa atau dana tambahan melalui layanan seperti GoPay.</p>

Setiap model menggambarkan setiap *topic*. dari *topic* di atas. Pada *topic* di atas terdapat beberapa kata yang sama namun berbeda nilai probabilitasnya. Seperti pada model topic 2, 4, 5, 6, 7, 9, dan 10 memiliki kata “akun” dengan nilai probabilitas berbeda-beda. Urutan berdasarkan pentingnya setiap *topic* terhadap *corpus*. Hal tersebut dapat dilihat dari Tabel 4.8

yang diurutkan berdasarkan nilai *marginal topic* yang diperoleh, seperti *topic-10* yang memiliki *marginal topic* sebesar 15,5% yang berisi topik sosial, *topic-6* sebesar 10.7% yang berisi topik *e-commerce*, *topic-1* sebesar 10.5% yang berisi topik sosial, *topic-2* sebesar 10.2% yang berisi topik platform media sosial, *topic-7* sebesar 9.8% yang berisi topik *e-commerce*, *topic-9* sebesar 9.1% yang berisi topik *e-commerce*, *topic-4* sebesar 9% yang berisi topik sosial, *topic-5* sebesar 8.9% yang berisi topik sosial, *topic-3* sebesar 8.3% yang berisi topik perbankan dan *topic 8* sebesar 8% yang berisi topik sosial.

#### 4.5 Iterasi 2

Pada iterasi ini, peneliti juga menambahkan percobaan menggunakan data baru yang digabung dari hasil *data scraping* yang dilakukan yang telah dijelaskan pada bab 4.1. Adanya percobaan ini untuk melihat bagaimana hasil akhir yang akan didapat dengan dataset yang lebih besar dari *dataset* sebelumnya. Dalam percobaan ini memiliki tahapan yang sama dengan percobaan yang telah dilakukan sebelumnya. Masih menggunakan teknik *data scraping* yang sama dengan perintah menggunakan *keyword* “penipuan” dengan rentang waktu selama tahun 2022. Data memiliki perbedaan pada jumlah *tweet* yang berhasil didapat. Gambar 4.24 memperlihatkan contoh data yang didapat.

	date	username	tweet
0	30/12/22	pucukubitawon	@TeamYouTube Saya tidak melakukan pelanggaran...
1	30/12/22	skincareygy	@ceemanizz Makasih tipsnyaa, berguna bgt supaya...
2	30/12/22	oyentidakbarbar	@ceemanizz Info nya penting bgt sih supaya sel...
3	30/12/22	biasabiasaw	@ceemanizz Lg marak bgt nih penipuan yg blg da...
4	30/12/22	cckmln	@ceemanizz Betul bgt! Skrg lg banyak bgt modus...
...	...	...	...
11576	12/12/22	daisy_smiley	3. Tiba tiba besoknya, 28 okt 2022. Semua cont...
11577	12/12/22	detik_sulsel	Setelah Wabup, Nama Bupati Pinrang Kini Dicatu...
11578	12/12/22	republikaonline	Usai lakukan penipuan modus arisan online, IRT...
11579	12/12/22	aigeaya	@gudyte ITU MAH PENIPUAN. Pernah diladenin sa...
11580	12/12/22	alilleen	@teuverse_id @treasuremembers Korban penipuan ...

Gambar 4.24 Dataset terbaru

Untuk meningkatkan keefektifan penelitian, peneliti menggabungkan kedua file menjadi satu, kemudian melakukan penghapusan data duplikat sebelum mengolahnya dan melihat hasilnya. Pada Gambar 4.25 menunjukkan implementasi penggabungan kedua *file* sehingga menghasilkan 18.882 data *tweet* yang akan diolah dari *dataset* pertama yang berjumlah 7.406 dan *dataset* kedua yang berjumlah 11.580.

```

def gabungkan_file_csv(file1, file2, file_tujuan):
    with open(file1, 'r', newline='') as f1:
        reader = csv.reader(f1)
        data1 = list(reader)

    with open(file2, 'r', newline='') as f2:
        reader = csv.reader(f2)
        data2 = list(reader)

    data_gabungan = data1 + data2

    with open(file_tujuan, 'w', newline='') as ft:
        writer = csv.writer(ft)
        writer.writerows(data_gabungan)

gabungkan_file_csv('s1922.csv', 'pn22.csv', 'gab.csv')

```

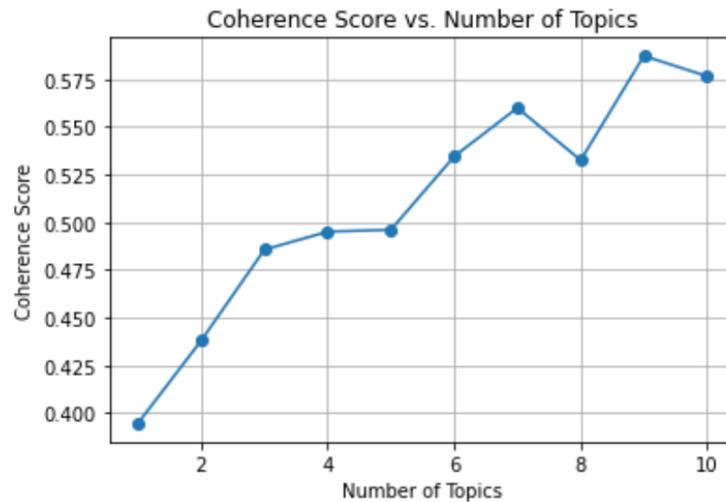
Gambar 4.25 Implementasi Menggabungkan dua *file*

	<b>date</b>	<b>username</b>	<b>tweet</b>
<b>0</b>	30/12/22	penipu_rakyat	[gak, bakal, kalo, rakyat, pinter, gak, bisa, ...
<b>1</b>	30/12/22	ssolsollaaa	[sedih, banget, akun, tiktok, gw, jadi, begini...
<b>2</b>	30/12/22	kieleily	[penipu, gaes, aku, ditipu, sama, dia]
<b>3</b>	30/12/22	lonjwinimine	[penipu, kontol, bangsat, gue, sampe, dibangun...
<b>4</b>	30/12/22	johyuzn	[help, rt, lagi, hatihati, guys, si, penipu, b...
...	...	...	...
<b>18878</b>	7/1/19	holiwigg	[aku, bner, gak, kuat, kalo, aku, di, posisi, ...
<b>18879</b>	7/1/19	tekukurmangap	[rezim, yg, paranoid, dengan, hoax, menunjukan...
<b>18880</b>	7/1/19	jelymonalisa	[coba, bayangin, mba, penipu, pun, ditipu, lua...
<b>18881</b>	6/1/19	nellydamayanti4	[jgnkan, org, lain, bs, dibohongi, alias, bs, ...
<b>18882</b>	5/1/19	masasetia	[mana, ada, orang, beriman, percaya, sama, pen...

Gambar 4.26 *Dataset* Gabungan

Untuk mengelola data gabungan terbaru dengan data lama memiliki persamaan proses, seperti terdapat *preprocessing* di dalamnya, yang membedakan hanyalah isi dari *stopwords* yang digunakan. Pada percobaan kedua ini, *stopwords* yang digunakan terdapat pada link berikut [https://drive.google.com/drive/folders/1aWQWa2mr6xvXL7FYeXkugxdpf-xwPRe0?usp=share\\_link](https://drive.google.com/drive/folders/1aWQWa2mr6xvXL7FYeXkugxdpf-xwPRe0?usp=share_link). Terdapat 3785 kata *stopwords* setelah dilakukannya pengujian model sebanyak 4 kali dari *stopwords* awal yang memiliki nilai 1637 kata. Penulis menambahkan beberapa kata seperti “ampun”, “bgst”, “dendam”, “kerana” dan lain sebagainya

yang penulis telah lampirkan pada bab lampiran. Dilakukannya pengujian model berulang dengan menambahkan kata pada stopwords untuk mendapatkan hasil yang maksimal dalam melakukan pemodelan topik. Berikut hasil dari percobaan kedua ini:



Gambar 4.27 Visualisasi grafik *Coherence Score*

Pada Gambar 4.27 memperlihatkan visualisasi grafik *Coherence Score* dan menunjukkan nilai tertinggi tertuju pada *topic* ke-9. Hal ini membuktikan jika jumlah *topic* yang terbentuk dari data terkait informasi penipuan di Twitter yang ada di Indonesia sebanyak 9 *topic*. Tabel 4.9 menunjukkan *coherence value* yang dimiliki oleh data kueri “penipuan”.

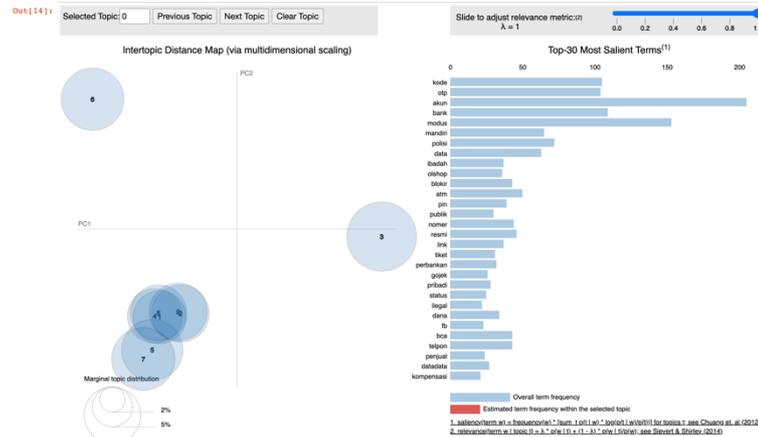
Tabel 4.9 *Coherence Value*

Num Topics	Coherence Value
1	0.394667
2	0.438007
3	0.485645
4	0.495067
5	0.496068
6	0.534461
7	0.559942
8	0.532523
<b>9</b>	<b>0.587142</b>
10	0.576502

Dalam model *topic modeling* yang telah dilakukan, terdapat nilai *coherence* tertinggi sebesar 0.587142 yang terkait dengan pengaturan jumlah topik sebanyak 9. Nilai ini menunjukkan bahwa model tersebut memiliki tingkat keterkaitan yang lebih baik antara kata-

kata dalam setiap topik yang dihasilkan. Berdasarkan hasil tersebut, dapat disimpulkan bahwa terdapat 9 topik yang membahas informasi penipuan yang ada pada Twitter di Indonesia.

Selanjutnya dilakukan analisis menggunakan LDA dengan melihat visualisasi *tools* pyLDAvis dan melihat keterkaitan antar kata suatu *topic*. Hasil LDA membuktikan jika pada 9 *topic* tersebut memiliki 30 *most term* sebagaimana pada Gambar 4.28.



Gambar 4.28 Visualisasi LDA dari Tweet “penipuan”

Pada visualisasi Gambar 4.28 terdapat *barplot* yang menampilkan frekuensi 30 kata yang sering muncul pada *topic*. Seperti yang telah dijelaskan sebelumnya, Gelembung besar menandakan topik yang paling populer di Twitter, sementara jarak dan tumpang tindih gelembung menunjukkan tingkat keterkaitan antar topik. Terlihat pada Gambar 4.28 memiliki gelembung yang bertabrakan sehingga dapat disimpulkan jika terdapat *topic* yang beririsan, yaitu:

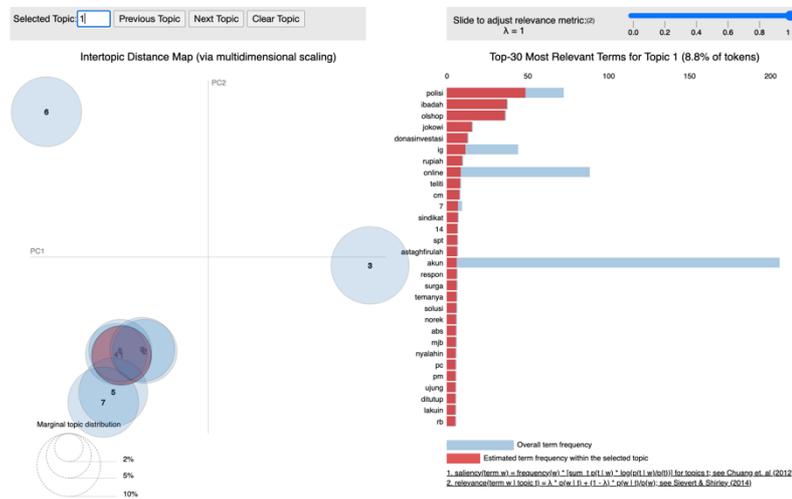
- *Topic 1* beririsan dengan *topic 2, 4, 5, 7, 8, dan 9*
- *Topic 2* beririsan dengan *topic 1, 4, 5, 7, 8 dan 9*
- *Topic 4* beririsan dengan *topic 1, 2, 5, 7, 8, dan 9*
- *Topic 5* beririsan dengan *topic 1, 2, 4, 7, 8, dan 9*
- *Topic 7* beririsan dengan *topic 1, 2, 4, 5, 8, dan 9*
- *Topic 8* beririsan dengan *topic 1, 2, 4, 5, 7, dan 9*
- *Topic 9* beririsan dengan *topic 1, 2, 4, 5, 7, dan 8*

Tabel 4.10 *Principal Component (PC)*

Kuadran 1	Kuadran 2	Kuadran 3	Kuadran 4
<i>Topic 3</i>	<i>Topic 3</i>	<i>Topic 6</i>	Irisan <i>Topic 1</i>
			Irisan <i>Topic 2</i>

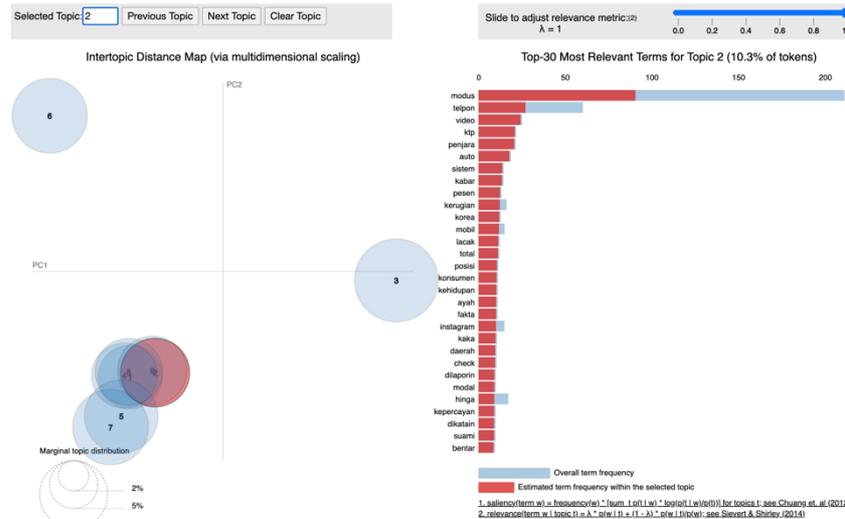
			Irisan <i>Topic 4</i>
			Irisan <i>Topic 5</i>
			Irisan <i>Topic 7</i>
			Irisan <i>Topic 8</i>
			Irisan <i>Topic 9</i>

Pada Gambar 4.28 menunjukkan gelembung yang tersebar pada empat *Principal Component* (PC) sebagaimana ditunjukkan pada pada Tabel 4.10. Masing-masing kuadran memiliki *topic* sendiri yang mana kuadran dua memiliki *topic* yang sama. Hal tersebut terlihat dari *topic* yang saling beririsan.

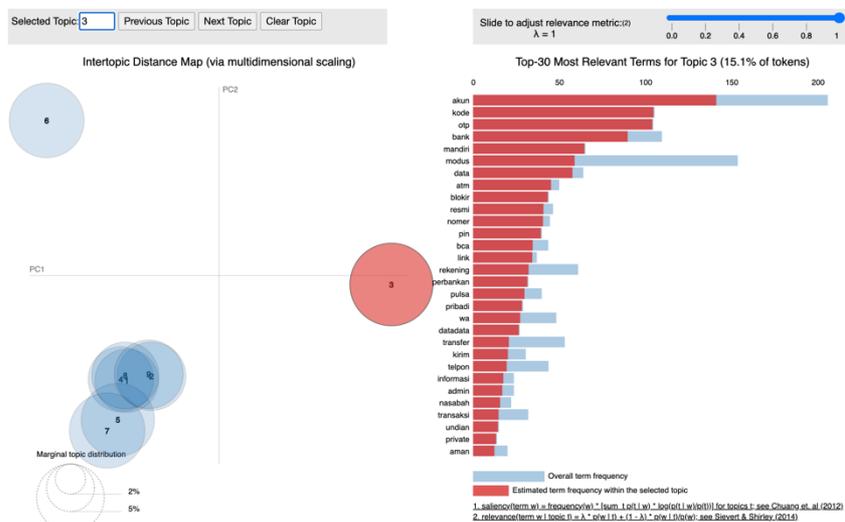


Gambar 4.29 Visualisasi *Topic-1*

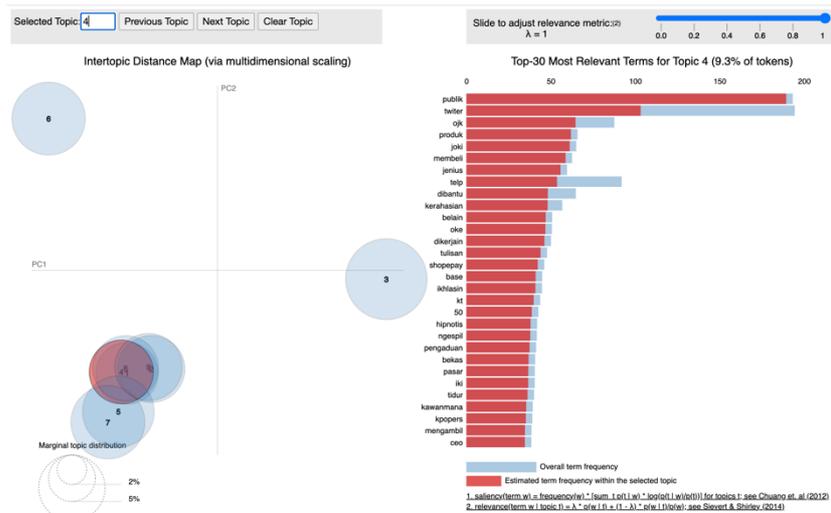
Kata yang muncul pada *topic* ke-1 terlihat pada Gambar 4.29. Dan kata-kata yang memiliki probabilitas tertinggi dan mewakili dari *topic* ke-1 yaitu  $0.015 * \text{"polisi"} + 0.012 * \text{"ibadah"} + 0.011 * \text{"olshop"} + 0.005 * \text{"jokowi"} + 0.004 * \text{"donasinvestasi"} + 0.004 * \text{"ig"} + 0.003 * \text{"rupiah"} + 0.003 * \text{"online"} + 0.003 * \text{"teliti"} + 0.003 * \text{"cm"}$ . Untuk nilai tertinggi terdapat pada *keyword* "polisi" dan "ibadah".

Gambar 4.30 Visualisasi *Topic-2*

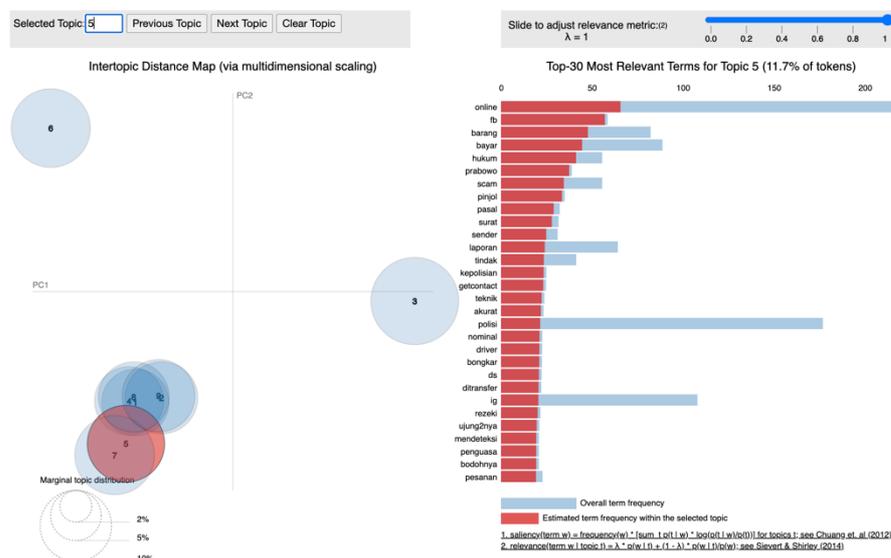
Kata yang muncul pada *topic* ke-2 terlihat pada Gambar 4.30. Dan kata-kata yang memiliki probabilitas tertinggi dan mewakili dari *topic* ke-2 yaitu :  $0.018 * \text{"modus"} + 0.005 * \text{"telpon"} + 0.005 * \text{"video"} + 0.004 * \text{"ktp"} + 0.004 * \text{"penjara"} + 0.004 * \text{"auto"} + 0.003 * \text{"sistem"} + 0.003 * \text{"kabar"} + 0.002 * \text{"pesen"} + 0.002 * \text{"kerugian"}$ . Untuk nilai tertinggi terdapat pada *keyword* "modus".

Gambar 4.31 Visualisasi *Topic-3*

Kata yang muncul pada *topic* ke-3 terlihat pada Gambar 4.31. Dan kata-kata yang memiliki probabilitas tertinggi dan mewakili dari *topic* ke-3 yaitu  $0.026 * \text{"akun"} + 0.019 * \text{"kode"} + 0.019 * \text{"otp"} + 0.017 * \text{"bank"} + 0.012 * \text{"mandiri"} + 0.011 * \text{"modus"} + 0.011 * \text{"data"} + 0.008 * \text{"atm"} + 0.008 * \text{"blokir"} + 0.008 * \text{"resmi"}$ . Untuk nilai tertinggi terdapat pada *keyword* "akun" dan "kode".

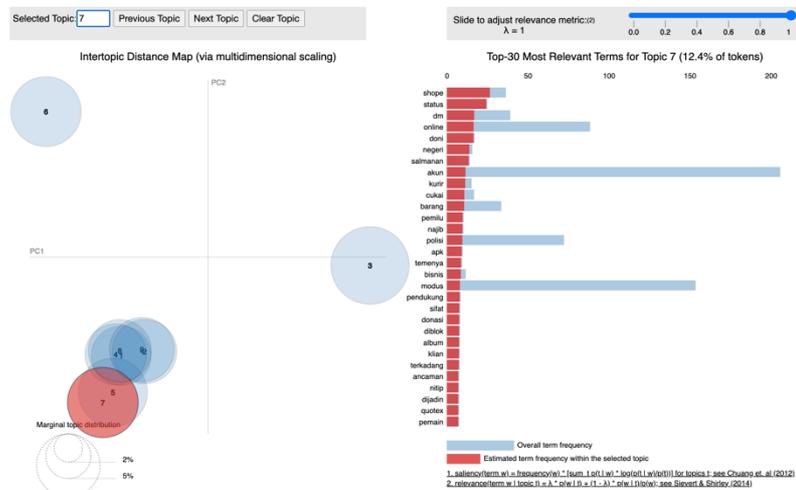
Gambar 4.32 Visualisasi *Topic-4*

Kata yang muncul pada *topic* ke-4 terlihat pada Gambar 4.32. Dan kata-kata yang memiliki probabilitas tertinggi dan mewakili dari *topic* ke-4 yaitu :  $0.009 * \text{"publik"} + 0.005 * \text{"twiter"} + 0.003 * \text{"ojk"} + 0.003 * \text{"produk"} + 0.003 * \text{"joki"} + 0.003 * \text{"membeli"} + 0.003 * \text{"jenius"} + 0.003 * \text{"telp"} + 0.002 * \text{"dibantu"} + 0.002 * \text{"kerahasian"}$ . Untuk nilai tertinggi terdapat pada *keyword* "publik" dan "twiter".

Gambar 4.33 Visualisasi *Topic-5*

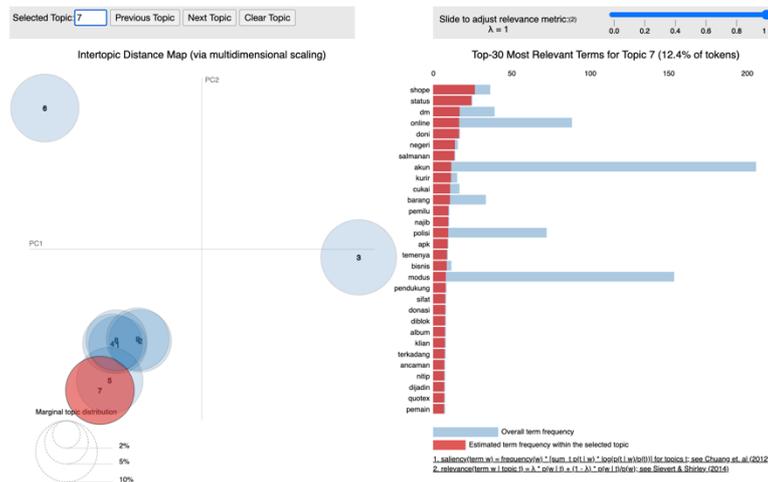
Kata yang muncul pada *topic* ke-5 terlihat pada Gambar 4.33. Dan kata-kata yang memiliki probabilitas tertinggi dan mewakili dari *topic* ke-5 yaitu :  $0.006 * \text{"online"} + 0.006 * \text{"fb"} + 0.005 * \text{"barang"} + 0.004 * \text{"bayar"} + 0.004 * \text{"hukum"} + 0.004 * \text{"prabowo"} + 0.003 * \text{"scam"} +$

$0.003 * \text{"pinjol"} + 0.003 * \text{"pasal"} + 0.003 * \text{"surat"}$ . Untuk nilai tertinggi terdapat pada *keyword* “online” dan “fb”.



Gambar 4.34 Visualisasi *Topic-6*

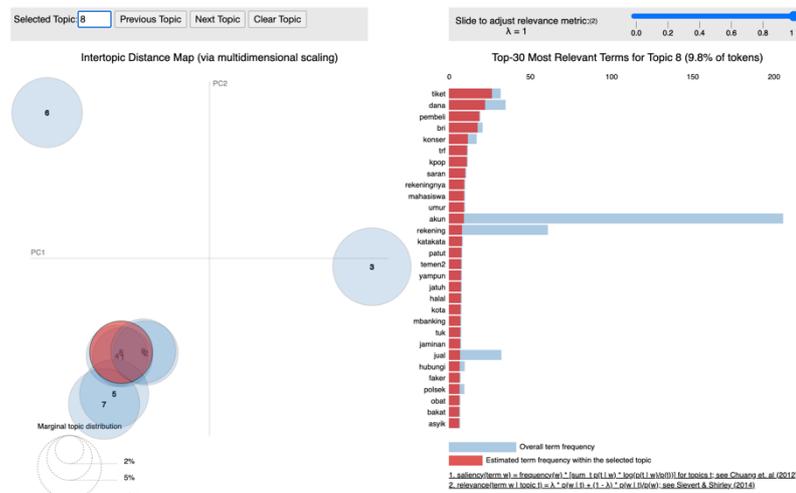
Kata yang muncul pada *topic* ke-6 terlihat pada Gambar 4.34. Dan kata-kata yang memiliki probabilitas tertinggi dan mewakili dari *topic* ke-6 yaitu :  $0.006 * \text{"gojek"} + 0.005 * \text{"penjual"} + 0.005 * \text{"film"} + 0.004 * \text{"gopay"} + 0.004 * \text{"saldo"} + 0.004 * \text{"akun"} + 0.003 * \text{"refund"} + 0.003 * \text{"politik"} + 0.003 * \text{"usaha"} + 0.003 * \text{"kantor"}$ . Untuk nilai tertinggi terdapat pada *keyword* “gojek”, “penjual”, dan “film”.



Gambar 4.35 Visualisasi *Topic-7*

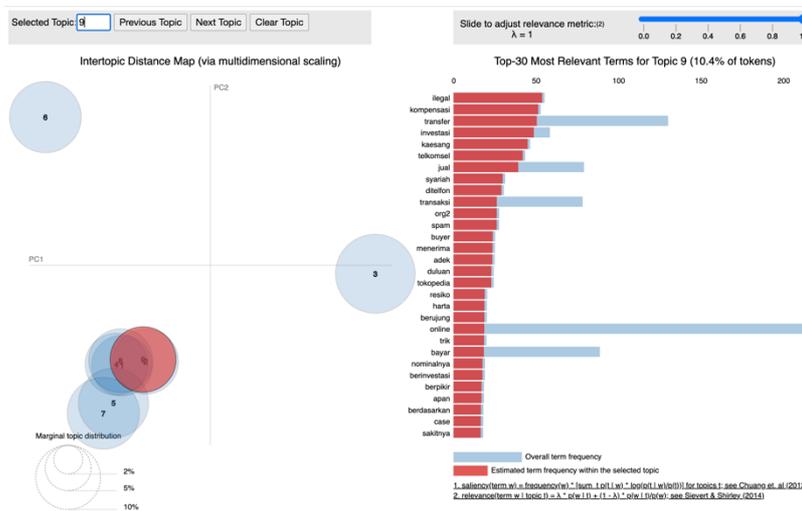
Kata yang muncul pada *topic* ke-7 terlihat pada Gambar 4.35. Dan kata-kata yang memiliki probabilitas tertinggi dan mewakili dari *topic* ke-7 yaitu  $0.006 * \text{"shope"} + 0.006 * \text{"status"} + 0.004 * \text{"dm"} + 0.004 * \text{"online"} + 0.004 * \text{"doni"} + 0.003 * \text{"negeri"} + 0.003 * \text{"salmanan"} +$

$0.003 * \text{"akun"} + 0.003 * \text{"kurir"} + "0.002 * \text{"cukai"}"$ . Untuk nilai tertinggi terdapat pada *keyword* “shope”, dan “status”.



Gambar 4.36 Visualisasi *Topic-8*

Kata yang muncul pada *topic* ke-8 terlihat pada Gambar 4.36. Dan kata-kata yang memiliki probabilitas tertinggi dan mewakili dari *topic* ke-8 yaitu :  $'0.008 * \text{"tiket"} + 0.006 * \text{"dana"} + 0.005 * \text{"pembeli"} + 0.005 * \text{"bri"} + "0.003 * \text{"konser"} + 0.003 * \text{"trf"} + 0.003 * \text{"kpop"} + 0.003 * \text{"saran"} + "0.003 * \text{"rekeningnya"} + 0.003 * \text{"mahasiswa}"$ . Untuk nilai tertinggi terdapat pada *keyword* “tiket” dan “dana”.



Gambar 4. 37 Visualisasi *Topic-9*

Kata yang muncul pada *topic* ke-9 terlihat pada Gambar 4.37. Dan kata-kata yang memiliki probabilitas tertinggi dan mewakili dari *topic* ke-9 yaitu :  $'0.006 * \text{"ilegal"} + 0.006 * \text{"kompensasi"} + 0.006 * \text{"transfer"} + 0.005 * \text{"investasi"} + 0.005 * \text{"kaesang"} +$

0.005\*"telkomsel" + 0.004\*"jual" + 0.003\*"syariah" + "0.003\*"ditelfon" + 0.003\*"transaksi".  
Dan untuk nilai tertinggi terdapat pada *keyword* “ilegal” “kompensasi” dan “transfer”.

Tabel 4.11 Hasil Perbandingan

Topic ke-	Model	Topic	Kesimpulan
1	'0.015*"polisi" + 0.012*"ibadah" + 0.011*"olshop" + 0.005*"jokowi" + "0.004*"donasinvestasi" + 0.004*"ig" + 0.003*"rupiah" + 0.003*"online" + "0.003*"teliti" + 0.003*"cm"	Sosial	Berdasarkan analisis pada data, model ini mengenai penipu yang menggunakan akun olshop di Instagram (IG) dengan modus penipuan donasi dan investasi palsu.
			
2	'0.018*"modus" + 0.005*"telpon" + 0.005*"video" + 0.004*"ktp" + "0.004*"penjara" + 0.004*"auto" + 0.003*"sistem" + 0.003*"kabar" + "0.002*"pesen" + 0.002*"kerugian"	Sosial	Berdasarkan analisis pada data, model ini mengenai modus yang umum seperti menelepon korban dan mengancam akan mengungkapkan video atau menggunakan informasi KTP mereka jika tidak memberikan uang.
			
3	'0.026*"akun" + 0.019*"kode" + 0.019*"otp" + 0.017*"bank" + 0.012*"mandiri" "+ 0.011*"modus" + 0.011*"data" + 0.008*"atm" + 0.008*"blokir" + "0.008*"resmi"	Perbankan	Berdasarkan analisis pada data, model ini mengenai orang-orang yang menjadi korban penipuan melalui modus yang melibatkan akun bank. Pelaku penipuan ini meminta kode OTP dari korban







## BAB V PENUTUP

### 5.1 Kesimpulan

Bagian ini akan merangkum kesimpulan yang diperoleh dari penelitian ini berdasarkan hasil analisis proses, algoritma, dan implementasi sistem yang telah dikembangkan. Selain itu, saran akan diberikan untuk mencatat dan memperbaiki aspek yang perlu ditingkatkan dalam penelitian ini di masa mendatang.

- a. Pengimplementasian metode *Latent Dirichlet Allocation* (LDA) pada penelitian dimulai dengan melakukan pengumpulan data, lalu dilanjutkan dengan preprocessing, melakukan pembobotan menggunakan TF-IDF, dan data siap dikelola menggunakan metode LDA. Dihasilkan 10 model dengan nilai koheren sebesar 0.600753 pada iterasi pertama dan pada iterasi kedua dihasilkan 9 model dengan nilai koheren sebesar 0.587142. Dilakukan percobaan kedua untuk melihat hasil pemodelan topik yang dilakukan dengan *dataset* yang lebih besar.
- b. Topik yang dimodelkan dari data Twitter selama kurun waktu 2019 hingga 2022 pada iterasi 1 yaitu 10 topik dan berisi 5 topik sosial, 3 topik *e-commerce*, 1 topik perbankan dan 1 topik platform media sosial. Sedangkan pada iterasi 2 yaitu terdapat 9 topik dan berisi 3 topik sosial, 2 topik *e-commerce*, 2 topik perbankan dan 2 topik platform media sosial.

### 5.2 Saran

Berdasarkan analisis mendalam yang dilakukan terhadap hasil dan pembahasan yang diperoleh dari penelitian ini, serta kesimpulan yang telah ditarik, disarankan untuk menerapkan strategi yang lebih proaktif dalam memperoleh hasil yang maksimal. Salah satu langkah yang dapat dilakukan adalah menambahkan dan memilih kata dari daftar *stopwords*. Dengan mengidentifikasi kata-kata yang kurang relevan sebagai *stopwords*, dapat membantu menganalisis hasil penelitian dengan lebih baik. Salah satu cara untuk melakukan ini yaitu dengan mempelajari daftar *stopwords* yang digunakan dan menambahkannya dengan kata-kata yang lebih spesifik dan sesuai konteks. Hal ini akan membantu memperbaiki kualitas analisis yang dilakukan dalam penelitian ini.

## DAFTAR PUSTAKA

- Arianto, B. W., & Anuraga, G. (2020). Pemodelan Topik Pengguna Twitter Mengenai Aplikasi “Ruangguru.” *Jurnal ILMU DASAR*, 21(2), 149–154.
- Bach, M. P., Krstič, Ž., Seljan, S., & Turulja, L. (2019). Text mining for big data analysis in financial sector: A literature review. *Sustainability (Switzerland)*, 11(5). <https://doi.org/10.3390/su11051277>
- Chilmi, M. L. C. (2021). *Latent Dirichlet Allocation (LDA) Untuk Mengetahui Topik Pembicaraan Warganet Twitter Tentang Omnibus Law* [Teknik Informatika]. UIN Syarif Hidayatullah Jakarta.
- Clinton, R. M. R. C., & Sengkey, R. (2019). Purwarupa Sistem Daftar Pelanggaran Lalulintas Berbasis Mini-Komputer Raspberry Pi. *Jurnal Teknik Elektro Dan Komputer*, 8(3).
- Dixon, S. (2022). *Countries with the most Twitter users 2022*. <https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/>
- ER, N. A. S. (2021). Implementasi Latent Dirichlet Allocation (LDA) Untuk Klasterisasi Cerita Berbahasa Bali. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 8(1), 127–134.
- Firdausi, A. (2022). *Kisah Logo Twitter Si Burung Biru, dari Manakah Nama Larry Bird Itu?* <https://tekno.tempo.co/read/1664100/kisah-logo-twitter-si-burung-biru-dari-manakah-nama-larry-bird-itu>
- Goyal, C. (2021). *Part 17: Step by Step Guide to Master NLP – Topic Modelling using pLSA*. <https://www.analyticsvidhya.com/blog/2021/06/part-17-step-by-step-guide-to-master-nlp-topic-modelling-using-plsa/>
- Gupta, H., & Patel, M. (2021). Method of Text Summarization Using Lsa and Sentence Based Topic Modelling with Bert. *Proceedings - International Conference on Artificial Intelligence and Smart Systems, ICAIS 2021*, 511–517. <https://doi.org/10.1109/ICAIS50930.2021.9395976>
- Hassan, S., Al-Augby, M., Mohammed, S. H., & Al-Augby, S. (2020). LSA & LDA Topic Modeling Classification: Comparison study on E-books Sentiment Analysis on Social Media View project LSA & LDA Topic Modeling Classification: Comparison study on E-books. *Indonesian Journal of Electrical Engineering and Computer Science*, 19(1). <https://doi.org/10.11591/ijeecs.v19.i1.pp%25p>

- Jayadianti, H., Damayanti, R., & Juwairiah. (2020). Latent Semantic Analysis (LSA) Dan Automatic Text Summarization (ATS) Dalam Optimasi Pencarian Artikel COVID 19. *Seminar Nasional Informatika 2020 (SEMNASIF 2020)*, 1(1).
- Khatulistiwa, G. (2021). *Heboh Penipuan Berkedok Donasi di Twitter, Ini Tips Hindari Tindak Kejahatan Ini di Internet*. <https://journal.sociolla.com/lifestyle/tips-hindari-penipuan-internet>
- Kompas.com. (2021). *Sejarah Twitter, Jejaring Sosial yang Terinspirasi dari SMS*. <https://tekno.kompas.com/read/2021/04/14/20420077/sejarah-twitter-jejaring-sosial-yang-terinspirasi-dari-sms?page=all>
- Kompas.com. (2022). *Mengenal Apa Itu Twitter dan Mengapa Orang Menggunakannya?* . <https://www.kompas.com/tren/read/2022/03/24/200500665/mengenal-apa-itu-twitter-dan-mengapa-orang-menggunakannya?page=all>
- Mandal, K. (2020). *Topic Modeling: Techniques and AI Models*. <https://dzone.com/articles/topic-modelling-techniques-and-ai-models>
- Manthiramoorathi, M. (2018). *Topic Modelling Techniques in NLP*. <https://iq.opengenius.org/topic-modelling-techniques/>
- Narendra, L. W. (2022). Topic Modeling in Conversational Dialogs for Naming Intent Labels Using LDA. *JTECS: Jurnal Sistem Telekomunikasi Elektronika Sistem Kontrol Power Sistem Dan Komputer*, 2(1), 65. <https://doi.org/10.32503/jtecs.v2i1.1820>
- Negara, E. S., Triadi, D., & Andryani, R. (2019). *Topic Modelling Twitter Data with Latent Dirichlet Allocation Method*.
- Nurlayli, A., & Nasichuddin, M. A. (2019). *Topic Modeling Penelitian Dosen JPTEI UNY pada Google Scholar Menggunakan Latent Dirichlet Allocation*. 4(2), 154–161. <https://doi.org/10.21831/elinvo.v4i2>
- Nursyafitri, G. D. (2022). *4 Alasan Data Scientist Gemar Menggunakan Python*.
- OOSGA. (2023). *The Social Indonesia - Social Media Outlook & Trends in 2023*.
- Polyzos, E., & Wang, F. (2022). Twitter and market efficiency in energy markets: Evidence using LDA clustered topic extraction. *Energy Economics*, 114. <https://doi.org/10.1016/j.eneco.2022.106264>
- Pryke, B. (2020, August 24). *How to Use Jupyter Notebook: A Beginner's Tutorial*. <https://www.dataquest.io/blog/jupyter-notebook-tutorial/>.
- Rashif, F., Ihza Perwira Nirvana, G., Alif Noor, M., & Aini Rakhmawati, N. (2021). Implementasi LDA untuk Pengelompokan Topik Cuitan Akun Bot Twitter bertagar

- #Covid-19 LDA Implementation for Topic of Bot's Tweets with #Covid-19 Hashtag. *Cogito Smart Journal* |, 7(1).
- Santoni, V., & Rufat, S. (2021). How fast is fast enough? Twitter usability during emergencies. *Geoforum*, 124, 20–35. <https://doi.org/10.1016/j.geoforum.2021.05.007>
- Sasmita, R. A., & Falani, A. Z. (2018). Pemanfaatan Algoritma TF/IDF Pada Sistem Informasi Ecomplaint Handling. *JURNAL LINK*, 27(1).
- Shalihah, N. F. (2021). *Ramai Penipuan "Cancel Order" di Twitter, Bagaimana Menyiasatinya?* <https://www.kompas.com/tren/read/2021/02/17/123200865/ramai-penipuan-cancel-order-di-twitter-bagaimana-menyiasatinya-?page=all>
- Shayne. (2021). *13 Types of Social Media You Should Be Using in 2022.* <https://invideo.io/blog/types-of-social-media/>
- Suhartono, D. (2018). *Latent Dirichlet Allocation (LDA).* <https://socs.binus.ac.id/2018/11/29/latent-dirichlet-allocation-lda/>
- Trinugraheni, N. F. (2022). *Waspada Modus Penipuan Terbaru di Twitter, Manfaatkan Kesuksesan NFT Moonbirds.* <https://www.tribunnews.com/techno/2022/04/19/waspada-modus-penipuan-terbaru-di-twitter-manfaatkan-kesuksesan-nft-moonbirds>
- Ulfa. (2022, February 9). *Mengenal Anaconda Dan Cara Meng-Install Anaconda Di Windows.* [Http://Labdas.Si.Fti.Unand.Ac.Id/2022/02/09/Mengenal-Anaconda-Dan-Cara-Menginstall-Anaconda-Di-Wondows/.](Http://Labdas.Si.Fti.Unand.Ac.Id/2022/02/09/Mengenal-Anaconda-Dan-Cara-Menginstall-Anaconda-Di-Wondows/)
- Yuliana W. (2023, January 18). *Apa Itu Docker? Pengertian Docker, Cara Kerja, dan Fungsinya.* [Https://Www.Hostinger.Co.Id/Tutorial/Apa-Itu-Docker.](Https://Www.Hostinger.Co.Id/Tutorial/Apa-Itu-Docker)
- Yuwita, N., Mauhibatillah, N., & Ulyah, H. '. (2022). DRAMATURGI: BUDAYA FLEXING BERKEDOK PENIPUAN DI MEDIA SOSIAL (STUDI KASUS INDRA KENZ DAN DONI SALMANAN). *Jurnal Komunikasi Dan Media*, 7(1).
- Zvornicanin, E. (2022, January 30). *Topic Modeling and Latent Dirichlet Allocation (LDA).* <https://datascienceplus.com/topic-modeling-and-latent-dirichlet-allocation-lda/>
- Zvornicanin, E. (2023). *When Coherence Score Is Good or Bad in Topic Modeling?* <https://www.baeldung.com/cs/topic-modeling-coherence-score#:~:text=There%20is%20no%20one%20way,want%20to%20maximize%20this%20score>

## LAMPIRAN

### a. Stopwords 357 kata

1	ada	47	beberapa	92	daripada	138	itukah	183	lain
2	adanya	48	seberapa	93	dekat	139	itulah	184	lainnya
3	adalah	49	begini	94	demi	140	jangan	185	melainkan
4	adapun	50	beginian	95	demikian	141	janganakan	186	selaku
5	agak	51	beginikah	96	demikianlah	142	janganlah	187	lalu
6	agaknya	52	beginilah	97	sedemikian	143	jika	188	melalui
7	agar	53	sebegini	98	dengan	144	jikalau	189	terlalu
8	akan	54	begitu	99	depan	145	juga	190	lama
9	akankah	55	begitukah	100	di	146	justru	191	lamanya
10	akhirnya	56	begitulah	101	dia	147	kala	192	selama
11	aku	57	begitupun	102	dialah	148	kalau	193	selama
12	akulah	58	sebegitu	103	dini	149	kalaulah	194	selamanya
13	amat	59	belum	104	diri	150	kalaupun	195	lebih
14	amatlah	60	belumah	105	dirinya	151	kalian	196	terlebih
15	anda	61	sebelum	106	terdiri	152	kami	197	bermacam
16	andalah	62	sebelumnya	107	dong	153	kamilah	198	macam
17	antar	63	sebenarnya	108	dulu	154	kamu	199	semacam
18	diantaranya	64	berapa	109	enggak	155	kamulah	200	maka
19	antara	65	berapakah	110	enggaknya	156	kan	201	makanya
20	diantaranya	66	berapalah	111	entah	157	kanan	202	makin
21	diantara	67	berapapun	112	entahlah	158	kapankah	203	malah
22	apa	68	betulkah	113	terhadap	159	kapapun	204	malahan
23	apaan	69	sebetulnya	114	terhadapnya	160	dikarenakan	205	mampu
24	mengapa	70	biasa	115	hal	161	karena	206	mampukah
25	apabila	71	biasanya	116	hampir	162	karenanya	207	mana
26	apakah	72	bila	117	hanya	163	ke	208	manakala
27	apalagi	73	bilakah	118	hanyalah	164	kecil	209	manalagi
28	apatah	74	bisa	119	harus	165	kemudian	210	masih
29	atau	75	bisakah	120	haruslah	166	kenapa	211	masihkah
30	ataukah	76	sebisanya	121	harusnya	167	kepada	212	semasih
31	ataupun	77	boleh	122	seharusnya	168	kepadanya	213	masing
32	bagai	78	bolehkah	123	hendak	169	ketika	214	mau
33	bagaikan	79	bolehlah	124	hendaklah	170	seketika	215	maupun
34	sebagai	80	buat	125	hendaknya	171	khususnya	216	semaunya
35	sebagainya	81	bukan	126	hingga	172	kini	217	memang
36	bagaimana	82	bukankah	127	sehingga	173	kinilah	218	mereka
37	bagaimanapun	83	bukanlah	128	ia	174	kiranya	219	merekalah
38	sebagaimana	84	bukannya	129	ialah	175	sekiranya	220	meski
39	bagaimanakah	85	cuma	130	ibarat	176	kita	221	meskipun
40	bagi	86	percuma	131	ingin	177	kitalah	222	semula
41	bahkan	87	dahulu	132	inginkah	178	kok	223	mungkin
42	bahwa	88	dalam	133	inginkan	179	lagi	224	mungkinkah
43	bahwasanya	89	dan	134	ini	180	lagian	225	nah
44	sebaliknya	90	dapat	135	inikah	181	selagi	226	namun
45	banyak	91	dari	136	inilah	182	lah	227	nanti
46	sebanyak	92	daripada	137	itu	183	lain	228	nantinya

229	nyaris	275	sedangkan		
230	oleh	276	sedikit		
231	olehnya	277	sedikitnya		
232	seorang	278	segala		
233	seseorang	279	segalanya		
234	pada	280	segera		
235	padanya	281	sesegera	320	suatu
236	padahal	282	sejak	321	sesudah
237	paling	283	sejenak	322	sesudahnya
238	sepanjang	284	sekali	323	sudah
239	pantas	285	sekalian	324	sudahkah
240	sepantasnya	286	sekalipun	325	sudahlah
241	sepantasnyalah	287	sesekali	326	supaya
242	para	288	sekaligus	327	tadi
243	pasti	289	sekarang	328	tadinya
244	pastilah	290	sekarang	329	tak
245	per	291	sekitar	330	tanpa
246	pernah	292	sekitarnya	331	setelah
247	pula	293	sela	332	telah
248	pun	294	selain	333	tentang
249	merupakan	295	selalu	334	tentu
250	rupanya	296	seluruh	335	tentulah
251	serupa	297	seluruhnya	336	tentunya
252	saat	298	semakin	337	tertentu
253	saatnya	299	sementara	338	seterusnya
254	sesaat	300	sempat	339	tapi
255	saja	301	semua	340	tetapi
256	sajalah	302	semuanya	341	setiap
257	saling	303	sendiri	342	tiap
258	bersama	304	sendirinya	343	setidaknya
259	sama	305	seolah	344	tidak
260	sesama	306	seperti	345	tidakkah
261	sambil	307	sepertinya	346	tidaklah
262	sampai	308	sering	347	toh
263	sana	309	seringnya	348	waduh
264	sangat	310	serta	349	wah
265	sangatlah	311	siapa	350	wahai
266	saya	312	siapakah	351	sewaktu
267	sayalah	313	siapapun	352	walau
268	se	314	disini	353	walaupun
269	sebab	315	disinilah	354	wong
270	sebabnya	316	sini	355	yaitu
271	sebuah	317	sinilah	356	yakni
272	tersebut	318	sesuatu	357	yang
273	tersebutlah	319	sesuatunya		
274	sedang	320	suatu		

b. Stopwords 1637 kata

1	papa	47	ente	93	agak	139	banyak	185	berlebihan
2	katakata	48	patut	94	agaknya	140	bapak	186	bermacam
3	mama	49	on	95	agar	141	baru	187	bermacam-macam
4	ade	50	ngancem	96	akan	142	bawah	188	bermaksud
5	kaka	51	emng	97	akankah	143	beberapa	189	bermula
6	istri	52	belain	98	akhir	144	begini	190	bersama
7	nyokap	53	keliatan	99	akhiri	145	beginian	191	bersama-sama
8	bokap	54	dikhianati	100	akhirnya	146	beginikah	192	bersiap
9	mami	55	berusaha	101	aku	147	beginilah	193	bersiap-siap
10	papi	56	yampun	102	akulah	148	begitu	194	bertanya
11	aduh	57	dpt	103	amat	149	begitukah	195	bertanya-tanya
12	pls	58	krna	104	amatlah	150	begitulah	196	berturut
13	tingi	59	bs	105	anda	151	begitupun	197	berturut-turut
14	nitip	60	bener2	106	andalah	152	bekerja	198	berturut
15	hubungan	61	curiga	107	antar	153	belakang	199	berujar
16	c	62	salam	108	antara	154	belakangan	200	berupa
17	nanya	63	klian	109	antaranya	155	belum	201	besar
18	ds	64	dibalik	110	apa	156	belumlah	202	betul
19	astaghfirullah	65	najib	111	apaan	157	benar	203	betulkah
20	kartu	66	get	112	apabila	158	benarkah	204	biasa
21	in	67	s	113	apakah	159	benarlah	205	biasanya
22	suaminya	68	suami	114	apalagi	160	berada	206	bila
23	kudu	69	nggeh	115	apatah	161	berakhir	207	bilakah
24	islam	70	membiarkan	116	artinya	162	berakhirlah	208	bisa
25	penipupenipu	71	kirain	117	asal	163	berakhirnya	209	bisakah
26	prabowo	72	ikhlasin	118	asalkan	164	berapa	210	boleh
27	jokowi	73	is	119	atas	165	berapakah	211	bolehkah
28	kaesang	74	ditinggalkan	120	atau	166	berapalah	212	bolehlah
29	lihat	75	tar	121	ataukah	167	berapapun	213	buat
30	pemerintah	76	palaku	122	ataupun	168	berarti	214	bukan
31	jemah	77	m	123	awal	169	berawal	215	bukankah
32	kapok	78	dibayar	124	awalnya	170	berbagai	216	bukanlah
33	ye	79	sial	125	bagai	171	berdatangan	217	bukannya
34	ngapain	80	pahala	126	bagaikan	172	beri	218	bulan
35	cuplikan	81	ae	127	bagaimana	173	berikan	219	bung
36	dikit	82	jenius	128	bagaimanakah	174	berikut	220	cara
37	sesuai	83	marah2	129	bagaimanapun	175	berikutnya	221	caranya
38	astaghfirullah	84	dituduh	130	bagi	176	berjumlah	222	cukup
39	ad	85	to	131	bagian	177	berkali-kali	223	cukupkah
40	surga	86	b	132	bahkan	178	berkata	224	cukuplah
41	sange	87	ratusan	133	bahwa	179	berkehendak	225	cuma
42	20	88	barang	134	bahwasanya	180	berkeinginan	226	dahulu
43	perasan	89	ada	135	baik	181	berkenaan	227	dalam
44	sj	90	adalah	136	bakal	182	berlainan	228	dan
45	folow	91	adanya	137	bakalan	183	berlalu	229	dapat
46	halah	92	adapun	138	balik	184	berlangsung	230	dari

231	daripada	277	dimintai	323	empat	369	jawabnya	415	kemungkinannya
232	datang	278	dimisalkan	324	enggak	370	jelas	416	kenapa
233	dekat	279	dimulai	325	enggaknya	371	jelaskan	417	kepada
234	demi	280	dimulailah	326	entah	372	jelastah	418	kepadanya
235	demikian	281	dimulainya	327	entahlah	373	jelasnya	419	kesampaian
236	demikianlah	282	dimungkinkan	328	guna	374	jika	420	keseluruhan
237	dengan	283	dini	329	gunakan	375	jikalau	421	keseluruhannya
238	depan	284	dipastikan	330	hal	376	juga	422	keterlaluan
239	di	285	diperbuat	331	hampir	377	jumlah	423	ketika
240	dia	286	diperbuatnya	332	hanya	378	jumlahnya	424	khususnya
241	diakhiri	287	dipergunakan	333	hanyalah	379	justru	425	kini
242	diakhirinya	288	diperkirakan	334	hari	380	kala	426	kinilah
243	dialah	289	diperlihatkan	335	harus	381	kalau	427	kira
244	diantara	290	diperlukan	336	haruslah	382	kalaupun	428	kira-kira
245	diantaranya	291	diperlukannya	337	harusnya	383	kalaupun	429	kiranya
246	diberi	292	dipersoalkan	338	hendak	384	kalian	430	kita
247	diberikan	293	dipertanyakan	339	hendaklah	385	kami	431	kitalah
248	diberikannya	294	dipunyai	340	hendaknya	386	kamilah	432	kok
249	dibuat	295	diri	341	hingga	387	kamu	433	kurang
250	dibuatnya	296	dirinya	342	ia	388	kamulah	434	lagi
251	didapat	297	disampaikan	343	ialah	389	kan	435	lagian
252	didatangkan	298	disebut	344	ibarat	390	kapankah	436	lah
253	digunakan	299	disebutkan	345	ibaratkan	391	kapankah	437	lain
254	diibaratkan	300	disebutkannya	346	ibaratnya	392	keapapun	438	lainnya
255	diibaratkannya	301	disini	347	ibu	393	karena	439	lalu
256	diingat	302	disinilah	348	ikut	394	karenanya	440	lama
257	diingatkan	303	ditambahkan	349	ingat	395	kasus	441	lamanya
258	diinginkan	304	ditandakan	350	ingat-ingat	396	kata	442	lanjut
259	dijawab	305	ditanya	351	ingin	397	katakan	443	lanjutnya
260	dijelaskan	306	ditanyai	352	inginkah	398	katakanlah	444	lebih
261	dijelaskannya	307	ditanyakan	353	inginkan	399	katanya	445	lewat
262	dikarenakan	308	ditegaskan	354	ini	400	ke	446	lima
263	dikatakan	309	ditujukan	355	inikah	401	keadaan	447	luar
264	dikatakannya	310	ditunjuk	356	inilah	402	kebetulan	448	macam
265	dikerjakan	311	ditunjuki	357	itu	403	kecil	449	maka
266	diketahui	312	ditunjukkan	358	itukah	404	kedua	450	makanya
267	diketahuinya	313	ditunjukkannya	359	itulah	405	keduanya	451	makin
268	dikira	314	ditunjuknya	360	jadi	406	keinginan	452	malah
269	dilakukan	315	dituturkan	361	jadilah	407	kelamaan	453	malahan
270	dilalui	316	dituturkannya	362	jadinya	408	kelihatan	454	mampu
271	dilihat	317	diucapkan	363	jangan	409	kelihatannya	455	mampukah
272	dimaksud	318	diucapkannya	364	jangankan	410	kelima	456	mana
273	dimaksudkan	319	diungkapkan	365	janganlah	411	keluar	457	manakala
274	dimaksudkannya	320	dong	366	jauh	412	kembali	458	manalagi
275	dimaksudnya	321	dua	367	jawab	413	kemudian	459	masa
276	diminta	322	dulu	368	jawaban	414	kemungkinan	460	masalah

461	masalahnya	507	menegaskan	523	mengucapkan	569	olehnya	615	sangat
462	masih	508	mengakhiri	524	mengucapkannya	570	pada	616	sangatlah
463	masihkah	509	mengapa	525	mengungkapkan	571	padahal	617	satu
464	masing	510	mengatakan	526	menjadi	572	padanya	618	saya
465	masing-masing	511	mengatakannya	527	menjawab	573	pak	619	sayalah
466	mau	512	mengenai	528	menjelaskan	574	paling	620	se
467	maupun	513	mengerjakan	529	menuju	575	panjang	621	sebab
468	melainkan	514	mengetahui	530	menunjuk	576	pantas	622	sebabnya
469	melakukan	515	menggunakan	531	menunjuki	577	para	623	sebagai
470	melalui	516	menghendaki	532	menunjukkan	578	pasti	624	sebagaimana
471	melihat	517	mengibaratkan	533	menunjuknya	579	pastilah	625	sebagainya
472	melihatnya	518	mengibaratkannya	534	menurut	580	penting	626	sebagian
473	memang	519	mengingat	535	menuturkan	581	pentingnya	627	sebaik
474	memastikan	520	mengingatkan	536	menyampaikan	582	per	628	sebaik-baiknya
475	memberi	521	menginginkan	537	menyangkut	583	percuma	629	sebaiknya
476	memberikan	522	mengira	538	menyatakan	584	perlu	630	sebaliknya
477	membuat	523	mengucapkan	539	menyebutkan	585	perlukah	631	sebanyak
478	memerlukan	524	mengucapkannya	540	menyeluruh	586	perlunya	632	sebegini
479	memihak	525	mengungkapkan	541	menyiapkan	587	pernah	633	sebegitu
480	meminta	526	menjadi	542	merasa	588	persoalan	634	sebelum
481	memintakan	527	menjawab	543	mereka	589	pertama	635	sebelumnya
482	memisalkan	528	menjelaskan	544	merekalah	590	pertama-tama	636	sebenarnya
483	memperbuat	529	menuju	545	merupakan	591	pertanyaan	637	seberapa
484	mempergunakan	530	menunjuk	546	meski	592	pertanyakan	638	sebesar
485	memperkirakan	531	menunjuki	547	meskipun	593	pihak	639	sebetulnya
486	memperlihatkan	532	menunjukkan	548	meyakini	594	pihaknya	640	sebisanya
487	mempersiapkan	533	menunjuknya	549	meyakinkan	595	pukul	641	sebuah
488	mempersoalkan	534	menurut	550	minta	596	pula	642	sebut
489	mempertanyakan	535	menuturkan	551	mirip	597	pun	643	sebutlah
490	mempunyai	536	menyampaikan	552	misal	598	punya	644	sebutnya
491	memulai	537	menyangkut	553	misalkan	599	rasa	645	secara
492	memungkinkan	538	menyatakan	554	misalnya	600	rasanya	646	secukupnya
493	menaiki	539	menyebutkan	555	mula	601	rata	647	sedang
494	menambahkan	540	menyeluruh	556	mulai	602	rupanya	648	sedangkan
495	menandaskan	541	menyiapkan	557	mulailah	603	saat	649	sedemikian
496	menanti	542	merasa	558	mulanya	604	saatnya	650	sedikit
497	menanti-nanti	543	mereka	559	mungkin	605	saja	651	sedikitnya
498	menantikan	544	merekalah	560	mungkinkah	606	sajalah	652	seenaknya
499	menanya	545	merupakan	561	nah	607	saling	653	segala
500	menanyai	546	meski	562	naik	608	sama	654	segalanya
501	menanyakan	547	meskipun	563	namun	609	sama-sama	655	segera
502	mendapat	548	meyakini	564	nanti	610	sambil	656	seharusnya
503	mendapatkan	549	meyakinkan	565	nantinya	611	sampai	657	sehingga
504	mendatang	550	minta	566	nyaris	612	sampai-sampai	658	seingat
505	mendatangi	551	mirip	567	nyatanya	613	sampaikan	659	sejak
506	mendatangkan	552	misal	568	oleh	614	sana	660	sejauh

661	sejenak	707	sendirinya	753	sudah	800	termasuk	846	yang	892	jgn
662	sejumlah	708	seolah	754	sudahkah	801	ternyata	847	kontol	893	no
663	sekadar	709	seolah-olah	755	sudahlah	802	tersampaikan	848	bangsat	894	deh
664	sekadarnya	710	seorang	756	supaya	803	tersebut	849	gak	895	jadi
665	sekali	711	sepanjang	757	tadi	804	tersebutlah	850	lg	896	ni
666	sekali-kali	712	sepantasnya	758	tadinya	805	tertentu	851	udh	897	mbak
667	sekalian	713	sepantasnyalah	759	tahu	806	tertuju	852	udah	898	sih
668	sekaligus	714	seperlunya	760	tahun	807	terus	853	si	899	trus
669	sekalipun	715	seperti	761	tak	808	terutama	854	ga	900	gitu
670	sekarang	716	sepertinya	762	tambah	809	tetap	855	ya	901	kayak
671	sekecil	717	sepihak	763	tambahnya	810	tetapi	856	yg	902	emang
672	seketika	718	sering	764	tampak	811	tiap	857	gw	903	untuk
673	sekiranya	719	seringnya	765	tampaknya	812	tiba	858	aja	904	tu
674	sekitar	720	serta	766	tandas	813	tiba-tiba	859	kak	905	trs
675	sekitarnya	721	serupa	767	tandasnya	814	tidak	860	gua	906	gt
676	sekurang-kurangnya	722	sesaat	768	tanpa	815	tidakkah	861	gue	907	langsung
677	sekurangnya	723	sesama	769	tanya	816	tidaklah	862	lu	908	gk
678	sela	724	sesampai	770	tanyakan	817	tiga	863	lo	909	juta
679	selagi	725	sesegea	771	tanyanya	818	tinggi	864	sm	910	gimana
680	selain	726	sesekali	772	tapi	819	toh	865	tuh	911	baru
681	selaku	727	seseorang	773	tegas	820	tunjuk	866	kg	912	hati
682	selalu	728	sesuatu	774	tegasnya	821	turut	867	tp	913	tdk
683	selama	729	sesuatunya	775	telah	822	tutor	868	amp	914	sampe
684	selama-lamanya	730	sesudah	776	tempat	823	tuturnya	869	nya	915	emg
685	selamanya	731	sesudahnya	777	tengah	824	ucap	870	ph	916	ditipu
686	selanjutnya	732	setelah	778	tentang	825	ucapnya	871	alah	917	penipu
687	seluruh	733	setempat	779	tentu	826	ujar	872	kena	918	tipu
688	seluruhnya	734	setengah	780	tentulah	827	ujarnya	873	pas	919	ketipu
689	semacam	735	seterusnya	781	tentunya	828	umum	874	bgt	920	menipu
690	semakin	736	setiap	782	tepat	829	umumnya	875	orang	921	nipu
691	semampu	737	setiba	783	terakhir	830	ungkap	876	org	922	tertipu
692	semampunya	738	setibanya	784	terasa	831	ungkapnya	877	w	923	penipuan
693	semasa	739	setidak-tidaknya	785	terbanyak	832	untuk	878	krn	924	tiputipu
694	semasih	740	setidaknya	786	terdahulu	833	usah	879	kalo	925	wkwkw
695	semata	741	setinggi	787	terdapat	834	usai	880	tau	926	anjir
696	semata-mata	742	seusai	788	terdiri	835	waduh	881	semoga	927	bpkibu
697	semaunya	743	sewaktu	789	terhadap	836	wah	882	klo	928	min
698	sementara	744	siap	790	terhadapnya	837	wahai	883	dgn	929	mah
699	semisal	745	siapa	791	teringat	838	waktu	884	biar	930	ati
700	semisalnya	746	siapakah	792	teringat-ingat	839	waktunya	885	utk	931	anjir
701	sempat	747	siapapun	793	terjadi	840	walau	886	jd	932	iya
702	semua	748	sini	794	terjadilah	841	walaupun	887	gabisa	933	hatihati
703	semuanya	749	sinilah	795	terjadinya	842	wong	888	nga	934	awas
704	semula	750	soal	796	terkira	843	yaitu	889	kau	935	nder
705	sendiri	751	soalnya	797	terlalu	844	yakin	890	bener	936	hati2
706	sendirian	752	suatu	798	terlebih	845	yakni	891	banget	937	sadar
				799	terlihat						

938	pinter	984	tiati	1030	di	1076	wk	1123	mendukung
939	dapet	985	ayo	1031	kebodohan	1077	drama	1124	dasar
940	guys	986	bla	1032	penyesalan	1078	wanita	1125	loh
941	bantu	987	kerjain	1033	rakyat	1079	mikir	1126	teriak
942	wkwk	988	tolak	1034	hahaha	1080	lawan	1127	pemimpin
943	bapakibu	989	mohon	1035	cebong	1081	seru	1128	mempertahankan
944	dikenal	990	bpk	1036	ngakak	1082	the	1129	tu
945	panik	991	merajalela	1037	senang	1083	bodoh	1130	dah
946	penuh	992	marak	1038	benci	1084	twist	1131	buktinya
947	waspada	993	ges	1039	ngerjain	1085	loh	1132	nonton
948	takut	994	gengs	1040	doi	1086	swindlers	1133	cerita
949	parah	995		1041	maling	1087	u	1134	bohong
950	jaman	996	berkedok	1042	emosi	1088	dgn	1135	lucu
951	kah	997	ngak	1043	goblok	1089	viral	1136	kasian
952	kalah	998	thread	1044	heran	1090	disuruh	1137	seneng
953	menolak	999	balik	1045	sekedar	1091	pengen	1138	beli
954	serem	1000	hai	1046	mahir	1092	cewek	1139	penipunya
955	takutnya	1001	namanya	1047	menyedari	1093	gampang	1140	mantan
956	minta	1002	terimakasih	1048	gerombolan	1094	cowo	1141	koq
957	yuk	1003	skrg	1049	nak	1095	dungu	1142	pandai
958	halo	1004	perasaan	1050	judulnya	1096	temenku	1143	polis
959	kebohongan	1005	terbaru	1051	mba	1097	percaya	1144	merasakan
960	berhatihati	1006	berani	1052	sdh	1098	jam	1145	bela
961	tiba	1007	beneran	1053	byk	1099	ikutan	1146	koalisi
962	gais	1008	dah	1054	pembohong	1100	barusan	1147	pura2
963	dunia	1009	skrng	1055	suka	1101	mudah	1148	kebohongan
964	memberikan	1010	wkwkwk	1056	takut	1102	besok	1149	nih
965	pintar	1011	berbagai	1057	ak	1103	parah	1150	abis
966	siap	1012	gini	1058		1104	mencegah	1151	eh
967	aneh	1013	tenang	1059	km	1105	bacalah	1152	kali
968	hidup	1014	mudah	1060	hah	1106	membohongi	1153	mo
969	sakit	1015	zaman	1061	calon	1107	lusa	1154	bego
970	basmi	1016	mencurigakan	1062	tukang	1108	ko	1155	malam
971	nih	1017	rawan	1063	udh	1109	diladeni	1156	haha
972	pokoknya	1018	sebar	1064	gitu	1110	gagal	1157	bangsat
973	kasih	1019	jaga	1065	bilang	1111	min	1158	gini
974	ngeri	1020	mintain	1066	salah	1112	salahkan	1159	yuk
975	yah	1021	haduh	1067	milih	1113	mas	1160	manusia
976	jelas	1022	suka	1068	pilih	1114	anak	1161	ngaku
977	bang	1023	aware	1069	kerajan	1115	penuh	1162	sang
978	puluhan	1024	cvcvcv	1070	marah	1116	ah	1163	kasihan
979	gaes	1025	sat	1071	dunia	1117	alias	1164	dibohongi
980	kudu	1026	ygy	1072	ceritanya	1118	makhluk	1165	korbanya
981	santai	1027	yg	1073	kah	1119	galak	1166	buku
982	korban	1028	penipu	1074	tuhan	1120	azab	1167	deh
983	fix	1029	dan	1075	ni	1121	tolol	1168	liat
						1122	2x		

1169	twist	1215	g	1261	bahasa	
1170	bro	1216	gapapa	1262	gmna	
1171	lupa	1217	melindungi	1263	drpd	
1172	wkwkwkw	1218	dimana	1264	p	
1173	wkwkwkwk	1219	mg	1265	ketawa	
1174	pusing	1220	dukung	1266	cek	
1175	ketemu	1221	susah	1267	movement	
1176	jugak	1222	rezim	1268	mahluk	
1177	twit	1223	mati	1269	ntar	
1178	bapaknya	1224	pintar	1270	cinta	
1179	menolak	1225	hadiah	1271	a	
1180	sy	1226	dh	1272	mah	
1181	kadrun	1227	melayu	1273	plot	
1182	situ	1228	tega	1274	bikin	
1183	aneh	1229	x	1275	ha	
1184	bagus	1230	i	1276	woi	
1185	kaum	1231	acount	1277	berteman	
1186	kemaren	1232	ama	1278	bosan	
1187	tg	1233	berhasil	1279	hahah	
1188	pasang	1234	je	1280	mari	
1189	negara	1235	la	1281	tingal	
1190	ngerasa	1236	malaysia	1282	ramai	
1191	nasib	1237	bos	1283	diraja	
1192	intinya	1238	ikhlas	1284	belajar	
1193	pakai	1239	kerja	1285	aing	
1194		24	1240	habis	1286	kisah
1195	bukanya	1241	beneran	1287	prnh	
1196	al	1242	setan	1288	tk	
1197	ka	1243	bangsa	1289	kakanya	
1198	dosa	1244	kalah	1290	sabar	
1199	dr	1245	selamat	1291	y	
1200	bawa	1246	mama	1292	gatau	
1201	kadang	1247	hahahaha	1293	menasihatkan	
1202	emak	1248	cepat	1294	berkalikali	
1203	muda	1249	ibuku	1295	makan	
1204	yo	1250	korban	1296	adik	
1205	biarkan	1251	jujur	1297	tungu	
1206	balasan	1252	kl	1298	nasihat	
1207	cuy	1253	bingung	1299	malu	
1208	cewe	1254	lainya	1300	dasarnya	
1209	ngomong	1255		5	1301	percakapan
1210	bersalah	1256	sayang	1302	sehat	
1211	tolong	1257	memuja	1303		0
1212	pagi	1258	kecewa	1304	kesian	
1213	membohongi	1259	esemka	1305	rame	
1214	d	1260	isi	1306	jaman	

1307	kerjanya	1353	mcm	1399	dg
1308	isinya	1354	akal	1400	sebenarnya
1309	n	1355	masuk	1401	tobat
1310	beda	1356	kecuali	1402	ilmu
1311	ku	1357	kek	1403	ditindak
1312	malem	1358	sbg	1404	suruh
1313	mengatasnamakan	1359	muka	1405	ketetapan
1314	main	1360	yah	1406	menang
1315	wajar	1361	mahu	1407	kayaknya
1316	keluarga	1362	ati	1408	gaes
1317	mampus	1363	selesai	1409	kosong
1318	hebat	1364	tobatnya	1410	pd
1319	mrk	1365	keren	1411	parpol
1320	wajar	1366	hm	1412	lha
1321	niat	1367	map	1413	anaknya
1322	indonesia	1368	it	1414	smp
1323	blnja	1369	banga	1415	bayangin
1324	janji	1370	presiden	1416	tipuan
1325	aktif	1371	terbukti	1417	capek
1326	lho	1372	mangaku	1418	12
1327	2019	1373	pake	1419	ribu
1328	licik	1374	paham	1420	ampun
1329	ketauan	1375	tmn	1421	bgst
1330	and	1376	kaget	1422	end
1331	diam	1377	vcs	1423	dendam
1332	kira2	1378	doang	1424	memilih
1333	masyarakat	1379	kelak	1425	kerana
1334	lives	1380	tonton	1426	inget
1335	anjing	1381	akhirat	1427	kejadian
1336	sehinga	1382	buta	1428	taunya
1337	suara	1383	sebel	1429	karma
1338	jutan	1384	ayat	1430	membela
1339	bukti	1385	orangorang	1431	bales
1340	dlm	1386	ketiga	1432	dibilang
1341	klu	1387		1433	endingnya
1342	penasaran	1388	yaudah	1434	belanja
1343	kwkwkw	1389	karna	1435	mengontrol
1344	lantas	1390	temen	1436	ahli
1345	rakyatnya	1391	teman	1437	ngakunya
1346	kuat	1392	butuh	1438	ditelpon
1347	menganggap	1393	hatihati	1439	sepatu
1348	telepon	1394	td	1440	khawatir
1349	aslinya	1395	tlp	1441	kakaknya
1350	kesel	1396	dirimu	1442	wajah
1351	berkuasa	1397	kaya	1443	nambah
1352	kesalahan	1398	kata2	1444	engkau







556	sebutnya	602	semakin	649	setiba	694	tepat	740	umum	786	bgt
557	secara	603	semampu	650	setibanya	695	terakhir	741	umumnya	787	orang
558	secukupnya	604	semampunya	651	setidak-tidaknya	696	terasa	742	ungkap	788	org
559	sedang	605	semasa	652	setidaknya	697	terbanyak	743	ungkapnya	789	w
560	sedangkan	606	semasih	653	setinggi	698	terdahulu	744	untuk	790	krn
561	sedemikian	607	semata	654	seusai	699	terdapat	745	usah	791	kalo
562	sedikit	608	semata-mata	655	sewaktu	700	terdiri	746	usai	792	tau
563	sedikitnya	609	semaunya	656	siap	701	terhadap	747	waduh	793	semoga
564	seenaknya	610	sementara	657	siapa	702	terhadapnya	748	wah	794	klo
565	segala	611	semisal	658	siapakah	703	teringat	749	wahai	795	dgn
566	segalanya	612	semisalnya	659	siapapun	704	teringat-ingat	750	waktu	796	biar
567	segera	613	sempat	660	sini	705	terjadi	751	waktunya	797	utk
568	seharusnya	614	semua	661	sinilah	706	terjadilah	752	walau	798	jd
569	sehingga	615	semuanya	662	soal	707	terjadinya	753	walaupun	799	gabisa
570	seingat	616	semula	663	soalnya	708	terkira	754	wong	800	nga
571	sejak	617	sendiri	664	suatu	709	terlalu	755	yaitu	801	kau
572	sejauh	618	sendirian	665	sudah	710	terlebih	756	yakin	802	bener
573	sejenak	619	sendirinya	666	sudahkah	711	terlihat	757	yakni	803	banget
574	sejumlah	620	seolah	667	sudahlah	712	termasuk	758	yang	804	jgn
575	sekadar	621	seolah-olah	668	supaya	713	ternyata	759	kontol	805	no
576	sekadarnya	622	seorang	669	tadi	714	tersampaikan	760	bangsat	806	deh
577	sekali	623	sepanjang	670	tadinya	715	tersebut	761	gak	807	jadi
578	sekali-kali	624	sepantasnya	671	tahu	716	tersebutlah	762	lg	808	ni
579	sekalian	625	sepantasnyalah	672	tahun	717	tertentu	763	udh	809	mbak
580	sekaligus	626	seperlunya	673	tak	718	tertuju	764	udah	810	sih
581	sekalipun	627	seperti	674	tambah	719	terus	765	si	811	trus
582	sekarang	628	sepertinya	675	tambahnya	720	terutama	766	ga	812	gitu
583	sekecil	629	sepihak	676	tampak	721	tetap	767	ya	813	kayak
584	seketika	630	sering	677	tampaknya	722	tetapi	768	yg	814	emang
585	sekiranya	631	seringnya	678	tandas	723	tiap	769	gw	815	untuk
586	sekitar	632	serta	679	tandasnya	724	tiba	770	aja	816	tu
587	sekitarnya	633	serupa	680	tanpa	725	tiba-tiba	771	kak	817	trs
588	sekurang-kurangnya	634	sesaat	681	tanya	726	tidak	772	gua	818	gt
589	sekurangnya	635	sesama	682	tanyakan	727	tidakkah	773	gue	819	langsung
590	sela	636	sesampai	683	tanyanya	728	tidaklah	774	lu	820	gk
591	selagi	637	sesegeera	684	tapi	729	tiga	775	lo	821	juta
592	selain	638	sesekali	685	tegas	730	tinggi	776	sm	822	gimana
593	selaku	639	seseorang	686	tegasnya	731	toh	777	tuh	823	baru
594	selalu	640	sesuatu	687	telah	732	tunjuk	778	jg	824	hati
595	selama	641	sesuatunya	688	tempat	733	turut	779	tp	825	tdk
596	selama-lamanya	642	sesudah	689	tengah	734	tutur	780	amp	826	sampe
597	selamanya	643	sesudahnya	690	tentang	735	tuturnya	781	nya	827	emg
598	selanjutnya	644	setelah	691	tentu	736	ucap	782	ph	828	ditipu
599	seluruh	645	setempat	692	tentulah	737	ucapnya	783	alah	829	penipu
600	seluruhnya	646	setengah	693	tentunya	738	ujar	784	kena	830	tipu
601	semacam	647	seterusnya	694	tepat	739	ujarnya	785	pas	831	ketipu
		648	setiap								

832	menipu	876	memberikan	922	wkwkwk	968	gitu	1014	as	1060	langkah
833	nipu	877	pintar	923	berbagai	969	deh	1015	ditemui	1061	aksi
834	tertipu	878	siap	924	gini	970	gais	1016	tuhan	1062	mengungkap
835	penipuan	879	aneh	925	tenang	971	kah	1017	ati	1063	mengalami
836	tiputipu	880	hidup	926	mudah	972	an	1018	temen	1064	penuh
837	wkwkw	881	sakit	927	zaman	973	ati2	1019	5	1065	makasih
838	anjir	882	basmi	928	mencurigakan	974	bilangnya	1020	unsur	1066	murah
839	bpkibu	883	nih	929	rawan	975	ges	1021	sigap	1067	habis
840	min	884	pokoknya	930	sebar	976	dr	1022	min	1068	ah
841	mah	885	kasih	931	jaga	977	62	1023	tpi	1069	hm
842	ati	886	ngeri	932	mintain	978	tu	1024	mah	1070	maju
843	anjir	887	yah	933	haduh	979	k	1025	yak	1071	kesel
844	iya	888	jelas	934	suka	980	bpkibu	1026	ka	1072	al
845	hatihati	889	bang	935	aware	981	sahabat	1027	t	1073	aset
846	awas	890	puluhan	936	cvcvcv	982	12	1028	indo	1074	ganti
847	nder	891	gaes	937	sat	983	maksa	1029	parah	1075	liat
848	hati2	892	kudu	938	yg	984	bodoh	1030	sayang	1076	jaman
849	sadar	893	santai	939	yg	985	i	1031	ngakunya	1077	kenal
850	pinter	894	korban	940	penipu	986	and	1032	please	1078	rawan
851	dapet	895	fix	941	dan	987	fe	1033	spil	1079	inget
852	guys	896	tiati	942	di	988	njir	1034	alhamdulillah	1080	bagus
853	bantu	897	ayo	943	yuk	989	to	1035	tenang	1081	enga
854	wkwk	898	bla	944	wkwkwkw	990	gila	1036	ramai	1082	bikin
855	bapakibu	899	kerjain	945	ot	991	kirain	1037	gmn	1083	mari
856	dikenal	900	tolak	946	dsb	992	we	1038	retwet	1084	buka
857	panik	901	mohon	947	anjing	993	sore	1039	dugan	1085	saranya
858	penuh	902	bpk	948	aduh	994	g	1040	terungkap	1086	untungnya
859	waspada	903	merajalela	949	hai	995	mah	1041	anak	1087	diwaspadai
860	takut	904	marak	950	halo	996	open	1042	selesai	1088	le
861	parah	905	ges	951	santai	997	enak	1043	terbesar	1089	cs
862	jaman	906	gengs	952	d	998	gada	1044	disuruh	1090	tolol
863	kah	907	1	953	ajg	999	2	1045	dasar	1091	livin
864	kalah	908	berkedok	954	rt	1000	ni	1046	gambar	1092	mengarahkan
865	menolak	909	ngak	955	tetep	1001	kek	1047	berguna	1093	mengunduh
866	serem	910	thread	956	km	1002	kali	1048	akibat	1094	tingkatkan
867	takutnya	911	balik	957	yah	1003	cek	1049	aman	1095	thanks
868	minta	912	hai	958	udh	1004	beneran	1050	kekya	1096	ngalamin
869	yuk	913	namanya	959	bnyc	1005	tolong	1051	ginian	1097	diakhir
870	halo	914	terimakasih	960	gini	1006	takut	1052	gata	1098	pastikan
871	kebohongan	915	skrg	961	nih	1007	80	1053	menghindari	1099	gercep
872	berhatihati	916	perasaan	962	kyk	1008	by	1054	10	1100	kondisi
873	tiba	917	terbaru	963	loh	1009	mantap	1055	a	1101	menerusi
874	gais	918	berani	964	4	1010	kendala	1056	nak	1102	fitnah
875	dunia	919	beneran	965	3	1011	kakak	1057	la	1103	tindakan
876	memberikan	920	dah	966	lakukan	1012	gasi	1058	the	1104	doang
877	pintar	921	skrng	967	menggunakan	1013	ak	1059	jalan	1105	maf

1106	terkait	1152	ih	1198	beda	1244	bacot	1290	background	1336	sialan
1107	ditangkap	1153	masi	1199	curiga	1245	enembe	1291	haha	1337	je
1108	mengirimkan	1154	dimanamana	1200	rep	1246	neraka	1292	engak	1338	msh
1109	help	1155	jdi	1201	upaya	1247	biadab	1293	random	1339	kaki
1110	mengatasnamakan	1156	34	1202	informasikan	1248	kalah	1294	raja	1340	ac
1111	membantu	1157	kmu	1203	menahan	1249	sy	1295	malu	1341	bakaran
1112	mengungkap	1158	warga	1204	kayaknya	1250	duh	1296	penyamun	1342	tai
1113	keren	1159	tujuan	1205	mr	1251	mu	1297	akalnya	1343	anjg
1114	aneh	1160	operandi	1206	meresahkan	1252	halah	1298	sumpah	1344	manusia
1115	perasan	1161	melindungi	1207	rela	1253	heran	1299	negara	1345	males
1116	salah	1162	terdakwa	1208	memperhambakan	1254	tergiur	1300	pakopang	1346	hahahaha
1117	berhasil	1163	takutnya	1209	kekerasan	1255	ngasih	1301	u	1347	komen
1118	lelang	1164	mangsa	1210	paksan	1256	eh	1302	blg	1348	ngaku
1119	202	1165	terbuka	1211	up	1257	bohong	1303	dp	1349	goblok
1120	rp	1166	pelangan	1212	baca	1258	ba	1304	keluarga	1350	jahat
1121	8	1167	pengumuman	1213	sosial	1259	bgst	1305	ngakak	1351	pn
1122	sdh	1168	jaga	1214	masuk	1260	ku	1306	kuat	1352	merugikan
1123	plis	1169	selamat	1215	kaya	1261	babi	1307	nama	1353	sya
1124	rame	1170	malam	1216	pake	1262	bego	1308	loe	1354	emosi
1125	alhamdulillah	1171	terhindar	1217	cal	1263	eh	1309	bg	1355	sialan
1126	dilaporkan	1172	kebohongan	1218	sobat	1264	ko	1310	id	1356	je
1127	isi	1173	waspadai	1219	menghubungi	1265	blm	1311	of	1357	msh
1128	biaya	1174	yuk	1220	mencurigakan	1266	lainya	1312	bau	1358	get
1129	ngirim	1175	bilang	1221	mending	1267	lebai	1313	wkwkwkwk	1359	ngerjain
1130	cepat	1176	marak	1222	gaes	1268	uname	1314	belum	1360	duit
1131	dijual	1177	serem	1223	share	1269	reply	1315	mf	1361	contact
1132	ujungnya	1178	masyarakat	1224	apapun	1270	ku	1316	teriak	1362	nyari
1133	foto	1179	kl	1225	ratusan	1271	lelaki	1317	kepala	1363	mati
1134	report	1180	petugas	1226	logi	1272	jelek	1318	akal	1364	dl
1135	ngirim	1181	namakan	1227	one	1273	gen	1319	sedih	1365	sebar
1136	dgn	1182	wajib	1228	ama	1274	ye	1320	blg	1366	karna
1137	kemaren	1183	mimin	1229	are	1275	wt	1321	ulung	1367	zaman
1138	harga	1184	belanja	1230	akhirakhir	1276	mengelabui	1322	kasian	1368	cangih
1139	biaya	1185	mas	1231	apresiasi	1277	be	1323	tai	1369	kartu
1140	twet	1186	b	1232	divonis	1278	huh	1324	anjg	1370	dikasih
1141	menjelang	1187	byk	1233	permintaan	1279	butuh	1325	manusia	1371	percaya
1142	bebas	1188	bener2	1234	dah	1280	hatiku	1326	males	1372	nomor
1143	viral	1189	gamau	1235	terburuk	1281	mata	1327	hahahaha	1373	gampang
1144	bebas	1190	bingung	1236	isu	1282	lho	1328	komen	1374	v
1145	semangat	1191	apa2	1237	metode	1283	jahanam	1329	ngaku	1375	testi
1146	berusaha	1192	abis	1238	modus2	1284	wtb	1330	goblok	1376	sory
1147	sarankan	1193	proses	1239	ciriciri	1285	wts	1331	jahat	1377	pergi
1148	sidang	1194	6	1240	diangkat	1286	pengen	1332	pn	1378	pakai
1149	blok	1195	orang2	1241	dikembalikan	1287	cat	1333	merugikan	1379	abad
1150	infoin	1196	hi	1242	bangsat	1288	arctic	1334	sya	1380	sok
1151	duitnya	1197	alami	1243	api	1289	monkeys	1335	emosi	1381	cantik



1705	aslinya	1751	teror	1797	omonngan	1843	r	1889	tlp	1935	grup	1981	kmm	2027	telepon
1706	diambil	1752	bales	1798	bisanya	1844	is	1890	viralkan	1936	in	1982	bernama	2028	jutan
1707	jenis	1753	diblokir	1799	taik	1845	bumi	1891	tl	1937	mingir	1983	busuk	2029	contoh
1708	dukung	1754	stay	1800	pinterpinter	1846	tandain	1892	sial	1938	dikirim	1984	lol	2030	trauma
1709	iklan	1755	tungu	1801	lgsg	1847	penipupenipu	1893	jenis	1939	jas	1985	mukanya	2031	bisanya
1710	menteri	1756	dianya	1802	pribadi	1848	info	1894	kat	1940	pelaku	1986	order	2032	suara
1711	kelas	1757	gtu	1803	tutup	1849	kilat	1895	sekolah	1941	tibatiba	1987	bls	2033	nyolong
1712	bong	1758	ma	1804	ha	1850	smart	1896	nyolong	1942	gatau	1988	masukin	2034	tulus
1713	secepat	1759	tarik	1805	bls	1851	penipupenipu	1897	anjeng	1943	modern	1989	dh	2035	diblock
1714	dunia	1760	marah	1806	undi	1852	fake	1898	dituduh	1944	menolak	1990	1504b	2036	talang
1715	kirim	1761	sumpahin	1807	nyesel	1853	pilih	1899	emak	1945	dimintain	1991	cerdas	2037	matin
1716	sgt	1762	login	1808	kata2	1854	rahasia	1900	post	1946	kelas	1992	aktif	2038	lahir
1717	angkat	1763	ngajak	1809	emak	1855	teror	1901	toko	1947	hadiah	1993	lakuin	2039	seneng
1718	pagi	1764	sempet	1810	ditelpon	1856	oknum	1902	angkat	1948	kasih	1994	dibayar	2040	nuduh
1719	dki	1765	menang	1811	kasih	1857	diambil	1903	seru	1949	9	1995	diblokir	2041	dipercaya
1720	jef	1766	video	1812	rek	1858	taik	1904	harganya	1950	jgan	1996	2024	2042	walaun
1721	sesungguhnya	1767	terbaik	1813	komplotan	1859	centang	1905	map	1951	sulit	1997	nyolong	2043	delete
1722	dihubungi	1768	rezeki	1814	barusan	1860	pilih	1906	dg	1952	rakyat	1998	manis	2044	chat
1723	lahir	1769	sial	1815	aps	1861	rahasia	1907	lihat	1953	dsr	1999	basa	2045	curang
1724	pilih	1770	base	1816	contoh	1862	block	1908	main	1954	intinya	2000	basi	2046	lumayan
1725	not	1771	centang	1817	terbaik	1863	ri	1909	survei	1955	secepat	2001	gonta	2047	tutup
1726	gajelas	1772	taunya	1818	pejabat	1864	dewan	1910	bodong	1956	delusion	2002	contoh	2048	omogan
1727	ti	1773	mintain	1819	ilmuwan	1865	identitas	1911	sejarah	1957	fenomena	2003	ky	2049	mengaku
1728	fakta	1774	kawan	1820	nomer	1866	sesungguhnya	1912	tim	1958	capgras	2004	jef	2050	takde
1729	mendukung	1775	resmi	1821	kumpulan	1867	sebenarnya	1913	jualan	1959	fakta	2005	jutan	2051	wk
1730	tobat	1776	tipu2	1822	bisanya	1868	sabar	1914	cuman	1960	penipunya	2006	dih	2052	editan
1731	sejati	1777	yng	1823	bales	1869	pagi	1915	mintain	1961	otak	2007	bahagia	2053	astaga
1732	talang	1778	bu	1824	tukang	1870	koruptor	1916	ha	1962	krna	2008	dianya	2054	penipuhati
1733	masak	1779	cvcvc	1825	profil	1871	yng	1917	jer	1963	ngutang	2009	dibales	2055	alfa
1734	dsr	1780	always	1826	nmr	1872	bocah	1918	doain	1964	tim	2010	satur	2056	rakyat
1735	bodong	1781	balasan	1827	mingir	1873	gede	1919	dicari	1965	masak	2011	bersih	2057	via
1736	15	1782	ikutan	1828	mendukung	1874	belajar	1920	always	1966	sengaja	2012	infonya	2058	disana
1737	mention	1783	uang	1829	in	1875	nyata	1921	stay	1967	berganti	2013	saking	2059	sekian
1738	diblock	1784	hadiah	1830	laporan	1876	mak	1922	sendernya	1968	dianggap	2014	kecuali	2060	ngira
1739	bocah	1785	penipunya	1831	trusted	1877	nyesel	1923	mreka	1969	aslinya	2015	kene	2061	jago
1740	ketauan	1786	pinter2	1832	berkedok	1878	lapor	1924	sempet	1970	sama2	2016	semalem	2062	carat
1741	mb	1787	kudu	1833	pemimpin	1879	10k	1925	keturunan	1971	sebel	2017	katain	2063	manipulatif
1742	real	1788	maling	1834	rakyat	1880	anus	1926	fuck	1972	besok	2018	bongin	2064	ikatan
1743	budak	1789	undian	1835	wa	1881	kudu	1927	sejati	1973	tlfn	2019	kapoknya	2065	malem
1744	adik	1790	viralkan	1836	maling	1882	taunya	1928	alasan	1974	klik	2020	jajanan	2066	haduh
1745	berganti	1791	komplotan	1837	trauma	1883	kemarin	1929	kawan	1975	terbaik	2021	bocil	2067	mikir
1746	delusion	1792	oh	1838	mendukung	1884	nelfon	1930	dont	1976	dicek	2022	hadiah	2068	nemu
1747	walaun	1793	vc	1839	otak	1885	ditelpon	1931	suci	1977	sesungguhnya	2023	telfon	2069	enaknya
1748	plus	1794	mukanya	1840	talang	1886	menang	1932	bertanggung	1978	marah	2024	nelpon	2070	sbg
1749	fenomena	1795	gangu	1841	150rb	1887	dunia	1933	retwetan	1979	anjink	2025	kasih	2071	dijuluki
1750	berkedok	1796	harganya	1842	9	1888	sms	1934	sungguh	1980	undi	2026	modusnya	2072	fuck



2350	biasanya	2396	didapat	2442	disampaikan	2488	ialah	2534	kan
2351	bila	2397	didatangkan	2443	disebut	2489	ibarat	2535	kapan
2352	bilakah	2398	digunakan	2444	disebutkan	2490	ibaratkan	2536	kapankah
2353	bisa	2399	diibaratkan	2445	disebutkannya	2491	ibaratnya	2537	kapanpun
2354	bisakah	2400	diibaratkannya	2446	disini	2492	ibu	2538	karena
2355	boleh	2401	diingat	2447	disinilah	2493	ikut	2539	karenanya
2356	bolehkah	2402	diingatkan	2448	ditambahkan	2494	ingat	2540	kasus
2357	bolehlah	2403	diinginkan	2449	ditandaskan	2495	ingat-ingat	2541	kata
2358	buat	2404	dijawab	2450	ditanya	2496	ingin	2542	katakan
2359	bukan	2405	dijelaskan	2451	ditanyai	2497	inginkah	2543	katakanlah
2360	bukankah	2406	dijelaskannya	2452	ditanyakan	2498	inginkan	2544	katanya
2361	bukanlah	2407	dikarenakan	2453	ditegaskan	2499	ini	2545	ke
2362	bukannya	2408	dikatakan	2454	ditujukan	2500	inikah	2546	keadaan
2363	bulan	2409	dikatakannya	2455	ditunjuk	2501	inilah	2547	kebetulan
2364	bung	2410	dikerjakan	2456	ditunjuki	2502	itu	2548	kecil
2365	cara	2411	diketahui	2457	ditunjukkan	2503	itukah	2549	kedua
2366	caranya	2412	diketuainya	2458	ditunjukkannya	2504	itulah	2550	keduanya
2367	cukup	2413	dikira	2459	ditunjuknya	2505	jadi	2551	keinginan
2368	cukupkah	2414	dilakukan	2460	dituturkan	2506	jadilah	2552	kelamaan
2369	cukuplah	2415	dilalui	2461	dituturkannya	2507	jadinya	2553	kelihatan
2370	cuma	2416	dilihat	2462	diucapkan	2508	jangan	2554	kelihatannya
2371	dahulu	2417	dimaksud	2463	diucapkannya	2509	jangankan	2555	kelima
2372	dalam	2418	dimaksudkan	2464	diungkapkan	2510	jangankanlah	2556	keluar
2373	dan	2419	dimaksudkannya	2465	dong	2511	jauh	2557	kembali
2374	dapat	2420	dimaksudnya	2466	dua	2512	jawab	2558	kemudian
2375	dari	2421	diminta	2467	dulu	2513	jawaban	2559	kemungkinan
2376	daripada	2422	dimintai	2468	empat	2514	jawabnya	2560	kemungkinannya
2377	datang	2423	dimisalkan	2469	enggak	2515	jelas	2561	kenapa
2378	dekat	2424	dimulai	2470	enggaknya	2516	jelaskan	2562	kepada
2379	demi	2425	dimulailah	2471	entah	2517	jelaslah	2563	kepadanya
2380	demikian	2426	dimulainya	2472	entahlah	2518	jelasnya	2564	kesampaian
2381	demikianlah	2427	dimungkinkan	2473	guna	2519	jika	2565	keseluruhan
2382	dengan	2428	dini	2474	gunakan	2520	jikalau	2566	keseluruhannya
2383	depan	2429	dipastikan	2475	hal	2521	juga	2567	keterlalu
2384	di	2430	diperbuat	2476	hampir	2522	jumlah	2568	ketika
2385	dia	2431	diperbuatnya	2477	hanya	2523	jumlahnya	2569	khususnya
2386	diakhiri	2432	dipergunakan	2478	hanyalah	2524	justru	2570	kini
2387	diakhirinya	2433	diperkirakan	2479	hari	2525	kala	2571	kinilah
2388	dialah	2434	diperlihatkan	2480	harus	2526	kalau	2572	kira
2389	diantara	2435	diperlukan	2481	haruslah	2527	kalaupun	2573	kira-kira
2390	diantaranya	2436	diperlukannya	2482	harusnya	2528	kalaupun	2574	kiranya
2391	diberi	2437	dipersoalkan	2483	hendak	2529	kalian	2575	kita
2392	diberikan	2438	dipertanyakan	2484	hendaklah	2530	kami	2576	kitalah
2393	diberikannya	2439	dipunyai	2485	hendaknya	2531	kamilah	2577	kok
2394	dibuat	2440	diri	2486	hingga	2532	kamu	2578	kurang
2395	dibuatnya	2441	dirinya	2487	ia	2533	kamulah	2579	lagi

2580	lagian	2627	memisalkan	2672	menjawab	2718	pak	2764	sayalah
2581	lah	2628	mempersuak	2673	menjelaskan	2719	paling	2765	se
2582	lain	2629	mempgunakan	2674	menuju	2720	panjang	2766	sebab
2583	lainnya	2630	mempkirakan	2675	menunjuk	2721	pantas	2767	sebabnya
2584	lalu	2631	memplihatkan	2676	menunjuki	2722	para	2768	sebagai
2585	lama	2632	mempersiapkan	2677	menunjukkan	2723	pasti	2769	sebagaimana
2586	lamanya	2633	mempersoalkan	2678	menunjuknya	2724	pastilah	2770	sebagainya
2587	lanjut	2634	mempertanyakan	2679	menurut	2725	penting	2771	sebagian
2588	lanjutnya	2635	mempunyai	2680	menuturkan	2726	pentingnya	2772	sebaik
2589	lebih	2636	memulai	2681	menyampaikan	2727	per	2773	sebaik-baiknya
2590	lewat	2637	memungkinkan	2682	menyangkut	2728	percuma	2774	sebaiknya
2591	lima	2638	menaiki	2683	menyatakan	2729	perlu	2775	sebaliknya
2592	luar	2639	menambahkan	2684	menyebutkan	2730	perlukah	2776	sebanyak
2593	macam	2640	menandaskan	2685	menyeluruh	2731	perlunya	2777	sebegini
2594	maka	2641	menanti	2686	menyiapkan	2732	pernah	2778	sebegitu
2595	makanya	2642	menanti-nanti	2687	merasa	2733	persoalan	2779	sebelum
2596	makin	2643	menantikan	2688	mereka	2734	pertama	2780	sebelumnya
2597	malah	2644	menanya	2689	mereka	2735	pertama-tama	2781	sebenarnya
2598	malahan	2644	menanya	2690	merupakan	2736	pertanyaan	2782	seberapa
2599	mampu	2645	menanyai	2691	meski	2737	pertanyakan	2783	sebesar
2600	mampukah	2646	menanyakan	2692	meskipun	2738	pihak	2784	sebetulnya
2601	mana	2647	mendapat	2693	meyakini	2739	pihaknya	2785	sebisanya
2602	manakala	2648	mendapatkan	2694	meyakinkan	2740	pukul	2786	sebuah
2603	manalagi	2649	mendatang	2695	minta	2741	pula	2787	sebut
2604	masa	2650	mendatangi	2696	mirip	2742	pun	2788	sebutlah
2605	masalah	2651	mendatangkan	2697	misal	2743	punya	2789	sebutnya
2606	masalahnya	2652	menegaskan	2698	misalkan	2744	rasa	2790	secara
2607	masih	2653	mengakhiri	2699	misalnya	2745	rasanya	2791	secukupnya
2608	masihkah	2654	mengapa	2700	mula	2746	rata	2792	sedang
2609	masing	2655	mengatakan	2701	mulai	2747	rupanya	2793	sedangkan
2610	masing-masing	2656	mengatakannya	2702	mulailah	2748	saat	2794	sedemikian
2611	mau	2657	mengenai	2703	mulanya	2749	saatnya	2795	sedikit
2612	maupun	2658	mengerjakan	2704	mungkin	2750	saja	2796	sedikitnya
2613	melainkan	2659	mengetahui	2705	mungkinkah	2751	sajalah	2797	seenaknya
2614	melakukan	2660	menggunakan	2706	nah	2752	saling	2798	segala
2615	melalui	2661	menghendaki	2707	naik	2753	sama	2799	segalanya
2616	melihat	2662	mengibaratkan	2708	namun	2754	sama-sama	2800	segera
2617	melihatnya	2663	mengibaratkannya	2709	nanti	2755	sambil	2801	seharusnya
2618	memang	2664	mengingat	2710	nantinya	2756	sampai	2802	sehingga
2619	memastikan	2665	mengingatkan	2711	nyaris	2757	sampai-sampai	2803	seingat
2620	memberi	2666	menginginkan	2712	nyatanya	2758	sampaikan	2804	sejak
2621	memberikan	2667	mengira	2713	oleh	2759	sana	2805	sejauh
2622	membuat	2668	mengucapkan	2714	olehnya	2760	sangat	2806	sejenak
2623	memerlukan	2669	mengucapkannya	2715	pada	2761	sangatlah	2807	sejumlah
2624	memihak	2670	mengungkapkan	2716	padahal	2762	satu	2808	sekadar
2625	meminta	2671	menjadi	2717	padanya	2763	saya	2809	sekadarnya
2626	memintakan								



3131	bla	3177	penyesalan	3223	wanita	3269	dasar	3315	bro	3362	melindungi	3408	drpd
3132	kerjain	3178	rakyat	3224	mikir	3270	loh	3316	lupa	3363	dimana	3409	p
3133	tolak	3179	hahaha	3225	lawan	3271	teriak	3317	wkwkwkw	3364	mg	3410	ketawa
3134	mohon	3180	cebong	3226	seru	3272	pemimpin	3318	wkwkwkwk	3365	dukung	3411	cek
3135	bpk	3181	ngakak	3227	the	3273	mempertahankan	3319	pusing	3366	susah	3412	movement
3136	merajalela	3182	senang	3228	bodoh	3274	tu	3320	ketemu	3367	rezim	3413	makhluk
3137	marak	3183	benci	3229	twist	3275	dah	3321	jugak	3368	mati	3414	ntar
3138	ges	3184	ngerjain	3230	loh	3276	buktinya	3322	twit	3369	pintar	3415	cinta
3139	gengs	3185	doi	3231	swindlers	3277	nonton	3323	bapaknya	3370	hadiah	3416	a
3140	1	3186	maling	3232	u	3278	cerita	3324	menolak	3371	dh	3417	mah
3141	berkedok	3187	emosi	3233	dgn	3279	bohong	3325	sy	3372	melayu	3418	plot
3142	ngak	3188	goblok	3234	viral	3280	lucu	3326	kadrun	3373	tega	3419	bikin
3143	thread	3189	heran	3235	disuruh	3281	kasian	3327	situ	3374	x	3420	ha
3144	balik	3190	sekedar	3236	pengen	3282	seneng	3328	aneh	3375	i	3421	woi
3145	hai	3191	mahir	3237	cewek	3283	beli	3329	bagus	3376	account	3422	berteman
3146	namanya	3192	nyedari	3238	gampang	3284	penipunya	3330	kaum	3377	ama	3423	bosan
3147	terimakasih	3193	gerombolan	3239	cowo	3285	mantan	3331	kemaren	3378	berhasil	3424	hahah
3148	skrg	3194	nak	3240	dungu	3286	koq	3332	tg	3379	je	3425	mari
3149	perasaan	3195	judulnya	3241	temenku	3287	pandai	3333	pasang	3380	la	3426	tingal
3150	terbaru	3196	mba	3242	percaya	3288	polis	3334	negara	3381	malaysia	3427	ramai
3151	berani	3197	sdh	3243	jam	3289	merasakan	3335	ngerasa	3382	bos	3428	diraja
3152	beneran	3198	byk	3244	ikutan	3290	bela	3336	nasib	3383	ikhlas	3429	belajar
3153	dah	3199	pembongong	3245	barusan	3291	koalisi	3337	intinya	3384	kerja	3430	aing
3154	skrng	3200	suka	3246	mudah	3292	pura2	3338	pakai	3385	habis	3431	kisah
3155	wkwkwk	3201	takut	3247	besok	3293	kebohongan	3339	24	3386	beneran	3432	prnh
3156	berbagai	3202	ak	3248	parah	3294	nih	3340	bukanya	3387	setan	3433	tk
3157	gini	3203	2	3249	mencegah	3295	abis	3341	al	3388	bangsa	3434	kakanya
3158	tenang	3204	km	3250	bacalah	3296	eh	3342	ka	3389	kalah	3435	sabar
3159	mudah	3205	hah	3251	membongong	3297	kali	3343	dosa	3390	selamat	3436	y
3160	zaman	3206	calon	3252	lusa	3298	mo	3344	dr	3391	mama	3437	gatau
3161	mencurigakan	3207	tukang	3253	ko	3299	bego	3345	bawa	3392	hahahahaha	3438	menasihatkan
3162	rawan	3208	udh	3254	diladeni	3300	malam	3346	kadang	3393	cepat	3439	berkalikali
3163	sebar	3209	gitu	3255	gagal	3301	haha	3347	emak	3394	ibuku	3440	makan
3164	jaga	3210	bilang	3256	min	3302	bangsat	3348	muda	3395	korban	3441	adik
3165	mintain	3211	salah	3257	salahkan	3303	gini	3349	yo	3396	jujur	3442	tungu
3166	haduh	3212	milih	3258	mas	3304	yuk	3350	biarkan	3397	kl	3443	nasihat
3167	suka	3213	pilih	3259	anak	3305	manusia	3351	balasan	3398	bingung	3444	malu
3168	aware	3214	kerajan	3260	penuh	3306	ngaku	3352	cuy	3399	lainya	3445	dasarnya
3169	cvcvcv	3215	marah	3261	ah	3307	sang	3353	cewe	3400	5	3446	percakapan
3170	sat	3216	dunia	3262	alias	3308	kasihan	3354	ngomong	3401	sayang	3447	sehat
3171	ygy	3217	ceritanya	3263	makhluk	3309	dibongong	3355	bersalah	3402	memuja	3448	0
3172	yg	3218	kah	3264	galak	3310	korbanya	3356	tolong	3403	kecewa	3449	kesian
3173	penipu	3219	tuhan	3265	azab	3311	buku	3357	pagi	3404	esemka	3450	rame
3174	dan	3220	ni	3266	tolol	3312	deh	3358	membongong	3405	isi	3451	jaman
3175	di	3221	wk	3267	2x	3313	liat	3359	d	3406	bahasa	3452	kerjanya
3176	kebodohan	3222	drama	3268	mendukung	3314	twist	3360	g	3407	gmna	3453	isinya
								3361	gaopaa				

