

**ANALISIS *TOPIC MODELLING* MENGENAI
PEMBERLAKUAN PEMBatasan KEGIATAN
MASYARAKAT MENGGUNAKAN *LATENT DIRECTIONAL
ALLOCATION (LDA)***

(Studi Kasus : *Tweet PPKM* di Media Sosial *Twitter*)

TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana Program
Studi Statistika



Disusun Oleh:

Desy Endriani

18611007

**PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS ISLAM INDONESIA
YOGYAKARTA
2022**

HALAMAN PERSETUJUAN PEMBIMBING TUGAS AKHIR

Judul : Analisis *Topic Modelling* Mengenai Pemberlakuan
Pembatasan Kegiatan Masyarakat Menggunakan *Latent
Dirichlet Allocation (LDA)*
(Studi Kasus : *Tweet* PPKM di Media Sosial
Twitter)

Nama Mahasiswa : Desy Endriani

NIM : 18611007

**TUGAS AKHIR INI TELAH DIPERIKSA DAN DISETUJUI UNTUK
DIUJIKAN**

Yogyakarta, 21 Januari 2022

Pembimbing

(Ayundyan Kesumawarti, S.Si., M.Si)

Mengetahui,
Ketua Prodi Statistika

(Dr. Edy Widodo, S.Si., M.Si.)

HALAMAN PENGESAHAN
TUGAS AKHIR

**ANALISIS *TOPIC MODELLING* MENGENAI
PEMBERLAKUAN PEMBATAAN KEGIATAN
MASYARAKAT MENGGUNAKAN *LATENT DIRECTIONAL***

ALLOCATION (LDA)

(Studi Kasus : *Tweet* PPKM di Media Sosial *Twitter*)

Nama Mahasiswa : Desy Endriani

NIM : 18611007

TUGAS AKHIR INI TELAH DIUJIKAN

PADA TANGGAL 11 Februari 2022 :

Nama Penguji

Tanda Tangan

1. Rahmadi Yotenka, M, Sc

.....

2. Arum Handini Primandari, S. Pd, Si, M, Si

.....

3. Ayundyah Kesumawati, S.Si., M.Si

.....

Mengetahui,

Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



(Prof. Riyanto, S.Pd., M.Si., Ph.D.)

KATA PENGANTAR



Assalamu'alaikum Wr.Wb

Puji syukur penulis panjatkan kepada Allah SWT, tuhan semesta alam yang telah memberikan kelancaran dalam menyelesaikan Tugas Akhir ini yang berjudul **“Analisis Topic Modelling Mengenai Pemberlakuan Pembatasan Kegiatan Masyarakat Menggunakan Latent Dirichlet Allocation (LDA)”**. Sholawat beserta salam semoga selamanya selalu terlimpah curahkan kepada baginda Muhammad SAW., kepada keluarganya, sahabatnya, serta umatnya yang senantiasa taat menjalankan perintahnya.

Tugas Akhir ini disusun untuk memenuhi persyaratan dalam menyelesaikan jenjang Strata Satu (S1) di Program Studi Statistika, Universitas Islam Indonesia. Dalam proses penyusunan Tugas Akhir ini, penulis mendapatkan banyak sekali bantuan, bimbingan serta dukungan dari berbagai pihak. Sehingga dalam kesempatan ini penulis bermaksud menyampaikan rasa terima kasih kepada:

1. Bapak Prof. Riyanto, S.Pd.,M.Si.,Ph.D., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Islam Indonesia.
2. Bapak Dr. Edy Widodo, M.Si., selaku Ketua Prodi Statistika, Universitas Islam Indonesia beserta jajarannya.
3. Ibu Ayundyah Kesumawati, S.Si., M.Si., selaku Dosen Pembimbing Tugas Akhir yang telah sabar bimbingan dan memberikan masukan, kepada penulis selama menyusun Tugas Akhir ini.
4. Bapak Enduy dan Ibu Sri Hernimawati selaku orang tua dari penulis, Usi Endriani dan Deni Jaenudin selaku kakak dan adik yang selalu memberikan semangat, doa, dan dukungannya.
5. Sahabat serta teman-teman seperjuangan Nadya Anis Faisa, Riefyal Arshyza Mustain, Diah Yuliyanti, Sefia Siti Afifah, Ridhwan Mustajab, dan Falah Novayanda yang selalu memberikan semangat untuk penulis menyelesaikan Tugas Akhir ini.

6. Teman-teman seperjuangan Statistika Angkatan 2018 yang senantiasa berjuang untuk memperoleh gelar S.Stat, dan terimakasih telah memberikan pengalaman yang tidak akan terlupakan selama kuliah.

Penulis menyadari bahwa Tugas Akhir ini masih jauh dari kata sempurna karena dengan segala keterbatasan yang ada. Untuk itu, penulis mengharapkan adanya kritik dan saran mengingat tidak ada sesuatu yang sempurna tanpa saran yang membangun. Harapan penulis semoga dengan Tugas Akhir ini dapat memberi manfaat bagi penulis dan pihak lain yang membacanya. Demikian Tugas Akhir ini dibuat, semoga dapat bermanfaat. Semoga Allah SWT selalu memberikan kebaikan dan ridha-Nya.

Wassalamualaikum Wr.Wb

Yogyakarta, 21 Januari 2022

A handwritten signature in black ink, appearing to read 'Desy Endriani', with a stylized flourish above it.

Desy Endriani

DAFTAR ISI

| | |
|--|------|
| HALAMAN JUDUL | i |
| HALAMAN PERSETUJUAN PEMBIMBING TUGAS AKHIR..... | ii |
| HALAMAN PENGESAHAN TUGAS AKHIR | iii |
| KATA PENGANTAR..... | iv |
| DAFTAR ISI | vi |
| DAFTAR TABEL | viii |
| DAFTAR GAMBAR..... | ix |
| DAFTAR LAMPIRAN | x |
| PERNYATAAN | xi |
| INTISARI..... | xii |
| ABSTRACT | xiii |
| BAB I PENDAHULUAN | 1 |
| 1.1. Latar Belakang Masalah..... | 1 |
| 1.2. Rumusan Masalah | 3 |
| 1.3. Batasan Masalah..... | 3 |
| 1.4. Tujuan Penelitian | 3 |
| 1.5. Manfaat Penelitian | 4 |
| BAB II TINJAUAN PUSTAKA | 5 |
| BAB III LANDASAN TEORI | 11 |
| 3.1. Kebijakan PPKM | 11 |
| 3.2. Media Sosial <i>Twitter</i> | 13 |
| 3.3. <i>Twitter Web Scrapping</i> | 14 |
| 3.4. <i>Machine Learning</i> | 14 |
| 3.5. <i>Data Mining</i> | 16 |
| 3.6. <i>Text Mining</i> | 16 |
| 3.7. Analisis Sentimen..... | 18 |
| 3.8. <i>Term Frequency - Inverse Document Frequency (TF-IDF)</i> | 19 |
| 3.9. <i>Natural Language Processing (NLP)</i> | 21 |
| 3.8.1 <i>Topic Modelling</i> | 22 |
| 3.8.2 <i>Latent Dirichlet Allocation (LDA)</i> | 22 |
| 3.8.3 <i>Gibbs Sampling</i> | 24 |
| 3.8.4 <i>Topic Coherence</i> | 25 |
| BAB IV METODOLOGI PENELITIAN..... | 27 |
| 4.1 Populasi dan Sampel | 27 |
| 4.2 Sumber Data..... | 27 |
| 4.3 Variabel dan Definisi Variabel..... | 27 |
| 4.4 Metode Pengambilan Data | 28 |
| 4.5 Metode Analisis Data..... | 28 |
| 4.6 Tahapan Penelitian | 28 |
| BAB V HASIL DAN PEMBAHASAN | 30 |
| 5.1 Analisis Deskriptif | 30 |
| 5.2 <i>Preprocessing</i> | 32 |
| 5.2.1 <i>Cleaning</i> | 32 |
| 5.2.2 <i>Case Folding</i> | 33 |
| 5.2.3 <i>Filtering</i> | 33 |

| | | |
|----------------------|--|----|
| 5.2.4 | <i>Tokenizing</i> | 34 |
| 5.2.5 | <i>Stemming</i> | 34 |
| 5.3 | Analisis Sentimen..... | 35 |
| 5.4 | <i>Term Frequency Invers Document Frequency (TF-IDF)</i> | 37 |
| 5.5 | Hasil <i>Topic Modelling</i> dengan Bigram Trigram | 40 |
| 5.6 | Hasil <i>Topic Modelling</i> dengan Unigram | 48 |
| BAB VI PENUTUP | | 57 |
| 6.1. | Kesimpulan | 57 |
| 6.2. | Saran..... | 58 |
| DAFTAR PUSTAKA..... | | 59 |
| LAMPIRAN | | 63 |

DAFTAR TABEL

| | |
|--|----|
| Tabel 2.1 Deskripsi Penelitian Terdahulu | 8 |
| Tabel 4.1 Variabel Penelitian | 27 |
| Tabel 5.1 Data PPKM | 30 |
| Tabel 5.2 Hasil <i>Cleaning</i> | 32 |
| Tabel 5.3 Hasil <i>Case Folding</i> | 33 |
| Tabel 5.4 Hasil <i>Filtering</i> | 34 |
| Tabel 5.5 Hasil <i>Tokenizing</i> | 34 |
| Tabel 5.6 Hasil <i>Stemming</i> | 35 |
| Tabel 5.7 Contoh daftar kata <i>lexicon</i> positif | 35 |
| Tabel 5.8 Contoh daftar kata <i>lexicon</i> negatif | 36 |
| Tabel 5.9 Contoh Perhitungan Sentimen Netral..... | 36 |
| Tabel 5.10 Contoh Perhitungan Sentimen Negatif..... | 36 |
| Tabel 5.11 Contoh Perhitungan Sentimen Positif | 36 |
| Tabel 5.12 Contoh Data Pembobotan <i>TF-IDF</i> | 38 |
| Tabel 5.13 Contoh Perhitungan <i>TF</i> | 38 |
| Tabel 5.14 Contoh Perhitungan <i>DF</i> | 39 |
| Tabel 5.15 Contoh Perhitungan <i>TF-IDF</i> | 39 |
| Tabel 5.16 Nilai Koherensi Topik PPKM 5 dengan Bigram Trigram | 41 |
| Tabel 5.17 Hasil Model Topik dengan Bigram Trigram..... | 47 |
| Tabel 5.18 Nilai Koherensi Topik PPKM dengan Unigram | 49 |
| Tabel 5.19 Hasil Model Topik 5 dengan Unigram..... | 55 |

DAFTAR GAMBAR

| | |
|--|----|
| Gambar 1.1 Grafik Pengguna <i>Twitter</i> di Dunia | 1 |
| Gambar 3.1 Representasi grafis dari LDA | 23 |
| Gambar 4.1 Diagram Alir Penelitian..... | 28 |
| Gambar 5.1 Diagram batang jumlah <i>tweet</i> | 31 |
| Gambar 5.2 Diagram batang jumlah kata..... | 31 |
| Gambar 5.3 Persentase sentimen data <i>tweet</i> PPKM..... | 37 |
| Gambar 5.4 Grafik <i>Coherence Score</i> Data PPKM dengan Bigram Trigram | 41 |
| Gambar 5.5 Visualisai umum <i>Topic Modelling</i> PPKM dengan Bigram Trigram..... | 42 |
| Gambar 5.6 <i>Word cloud</i> Topik 1 dengan Bigram Trigram | 43 |
| Gambar 5.7 Visualisasi Topik 1 dengan Bigram Trigram | 44 |
| Gambar 5.8 Visualisasi Topik 2 dengan Bigram Trigram | 45 |
| Gambar 5.9 Visualisasi Topik 3 dengan Bigram Trigram | 45 |
| Gambar 5.10 Visualisasi Topik 4 dengan Bigram Trigram | 46 |
| Gambar 5.11 Visualisasi Topik 5 dengan Bigram Trigram | 47 |
| Gambar 5.12 Grafik <i>Coherence Score</i> Data PPKM dengan Unigram | 49 |
| Gambar 5.13 Visualisai umum <i>Topic Modelling</i> PPKM dengan Unigram..... | 50 |
| Gambar 5.14 <i>Word cloud</i> Topik 1 dengan Unigram | 51 |
| Gambar 5.15 Visualisasi Topik 1 dengan Unigram | 52 |
| Gambar 5.16 Visualisasi Topik 2 dengan Unigram | 53 |
| Gambar 5.17 Visualisasi Topik 3 dengan Unigram | 53 |
| Gambar 5.18 Visualisasi Topik 4 dengan Unigram | 54 |
| Gambar 5.19 Visualisasi Topik 5 dengan Unigram | 55 |

DAFTAR LAMPIRAN

| | |
|---|----|
| Lampiran 1 Data <i>Tweet</i> PPKM | 63 |
| Lampiran 2 <i>Script Scrapping</i> | 63 |
| Lampiran 3 <i>Script Grafik</i> | 63 |
| Lampiran 4 <i>Script Analisis Sentimen</i> | 63 |
| Lampiran 5 <i>Script TF-IDF</i> | 63 |
| Lampiran 6 <i>Script Topic Modelling</i> | 63 |
| Lampiran 7 <i>Word cloud</i> Topik 2 PPKM dengan Bigram Trigram | 63 |
| Lampiran 8 <i>Word cloud</i> Topik 3 PPKM dengan Bigram Trigram | 64 |
| Lampiran 9 <i>Word cloud</i> Topik 4 PPKM dengan Bigram Trigram | 64 |
| Lampiran 10 <i>Word cloud</i> Topik 5 PPKM dengan Bigram Trigram | 64 |
| Lampiran 11 <i>Word cloud</i> Topik 2 PPKM dengan Unigram | 65 |
| Lampiran 12 <i>Word cloud</i> Topik 3 PPKM dengan Unigram | 65 |
| Lampiran 13 <i>Word cloud</i> Topik 4 PPKM dengan Unigram | 65 |
| Lampiran 14 <i>Word cloud</i> Topik 4 PPKM dengan Unigram | 66 |

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya karya yang sebelumnya pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan orang lain, kecuali yang diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 21 Januari 2022



Desy Endriani

INTISARI

ANALISIS *TOPIC MODELLING* MENGENAI PEMBERLAKUAN PEMBATASAN KEGIATAN MASYARAKAT MENGGUNAKAN *LATENT DIRECHLET ALLOCATION (LDA)*

(Studi Kasus : *Tweet* PPKM di Media Sosial *Twitter*)

Desy Endriani

Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Islam Indonesia

Media sosial merupakan sarana yang memfasilitasi berbagi ide, pemikiran, dan informasi melalui jaringan dan komunitas virtual. Salah satu media sosial yang banyak digunakan masyarakat adalah *Twitter*. *Twitter* memberikan layanan jejaring sosial di mana pengguna dapat berbagi dan berinteraksi dengan pesan yang dikenal sebagai "*tweet*". Melalui sosial media *Twitter* ini peneliti ingin mengetahui perspektif masyarakat terhadap kebijakan PPKM atau Pemberlakuan Pembatasan Kegiatan Masyarakat. Kebijakan PPKM merupakan kebijakan pemerintah dalam upaya menekan penyebaran wabah covid-19. Pembatasan yang dimaksud yaitu pembatasan kegiatan yang biasa dilakukan oleh masyarakat dengan pengawasan setiap daerah yang memberlakukannya. Dengan adanya kebijakan ini, pemerintah berharap dapat meminimalisir aktivitas yang dapat memperluas penyebaran covid-19. Analisis sentiemen menunjukkan nilai persentase terbesar pada sentimen negatif yaitu sebesar 51,1%, yang artinya komentar tentang PPKM di *Twitter* kebanyakan masyarakat tidak mendukung ataupun menolak tentang adanya kebijakan ini. Menggunakan metode *Latent Dirichlet Allocation* untuk pemodelan topik perspektif masyarakat diperoleh hasil penelitian dengan bigram trigram sebanyak 5 topik tentang PPKM yaitu instruksi PPKM mikro, surat edar perpanjangan PPKM, protokol kesehatan saat PPKM mikro, PPKM Jawa Bali, dan PPKM darurat. Sedangkan untuk hasil penelitian dengan unigram sebanyak 5 topik juga yaitu protokol kesehatan, PPKM Jawa Bali, PPKM mikro, perpanjangan PPKM, dan surat edaran PPKM.

Kata Kunci : Pemberlakuan Pembatasan Kegiatan Masyarakat, PPKM, *Topic Modelling Latent Dirichlet Allocation*.

ABSTRACT

TOPIC MODELLING ANALYSIS OF THE IMPLEMENTATION OF COMMUNITY ACTIVITIES RESTRICTIONS USING LATENT DIRECHLET ALLOCATION (LDA)

(Case Study: PPKM Tweets on Twitter Social Media)

Desy Endriani

Department of Statistics, Faculty of Matematics and Natural Sciences
Universitas Islam Indonesia

Social media is a means that facilitates the sharing of ideas, thoughts, and information through virtual networks and communities. One of the most widely used social media is Twitter. Twitter provides a social networking service where users can share and interact with messages known as "tweets". Through social media Twitter, researchers want to know the community's perspective on the PPKM policy or the Enforcement of Restrictions on Community Activities. The PPKM policy is a government policy in an effort to suppress the spread of the COVID-19 outbreak. The restrictions in question are restrictions on activities that are usually carried out by the community with the supervision of each region that enforces them. With this policy, the government hopes to minimize activities that can expand the spread of covid -19. Sentiment analysis shows that the largest percentage value on negative sentiment is 51.1%, which means that comments about PPKM on Twitter most people do not support or reject this policy. Using the Latent Dirichlet Allocation method for modeling community perspective topics, the results obtained with a bigram trigram of 5 topics on PPKM, namely micro PPKM instructions, PPKM extension circulars, health protocols during micro PPKM, Java Bali PPKM, and emergency PPKM. As for the results of the study with unigram as many as 5 topics, namely health protocols, PPKM Java Bali, micro PPKM, extend PPKM, and PPKM circulars letters.

Keywords: *Enforcement of Restrictions on Community Activities, PPKM, Topic Modelling Latent Dirichlet Allocation.*

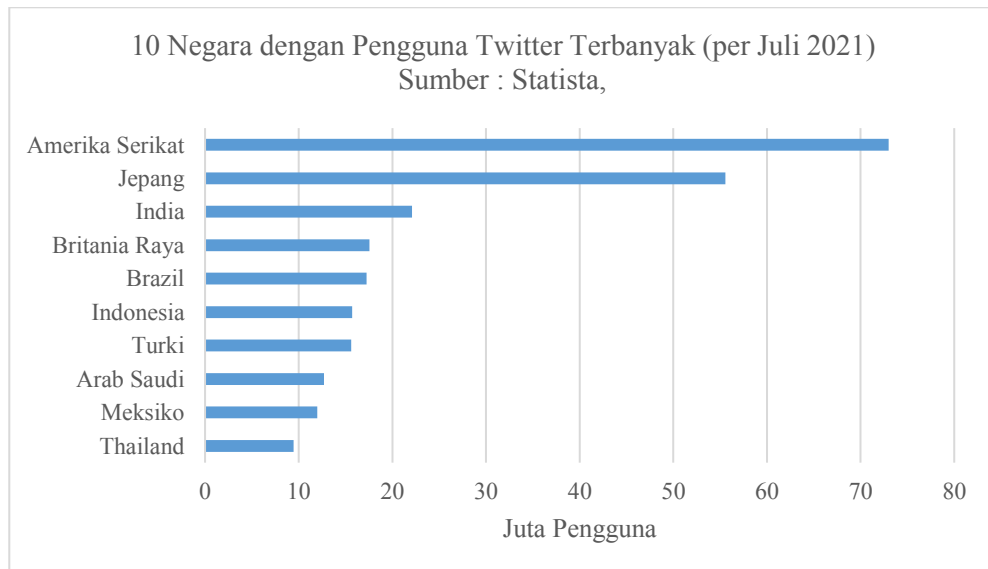
BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Media sosial adalah teknologi berbasis komputer yang memfasilitasi berbagai ide, pemikiran, dan informasi melalui pembangunan jaringan dan komunitas virtual. Kekuatan media sosial adalah kemampuan untuk terhubung dan berbagi informasi dengan siapa saja secara bersamaan. Analisis menunjukkan bahwa ada 4,55 miliar pengguna media sosial di seluruh dunia pada Oktober 2021, setara dengan 57,6% dari total populasi global. Jaringan media sosial terbesar termasuk *Facebook*, *YouTube*, *Instagram*, *Twitter*, dan *TikTok*. Media sosial biasanya menampilkan konten yang dibuat pengguna dan profil yang dipersonalisasi (*Global Social Media Stats*, 2021)

Twitter adalah salah satu media sosial paling populer di dunia. Banyak orang yang memanfaatkan platform ini di berbagai negara untuk berkomunikasi atau mendapatkan informasi termasuk Indonesia. Berikut data yang menunjukkan bahwa Indonesia termasuk dalam 10 negara dengan pengguna *Twitter* terbesar di dunia :



Gambar 1.1 Grafik Pengguna *Twitter* di Dunia
(databoks, 2021)

Grafik pada **Gambar 1.1** menunjukkan Indonesia berada di peringkat ke 6 untuk kategori pengguna *Twitter* terbanyak di dunia yaitu sebanyak 15,7 juta pengguna. *Twitter* merupakan situs berita dan jejaring sosial *online* tempat dimana

seseorang dapat berkomunikasi dalam pesan singkat yang disebut *tweet*. *Tweeting* adalah memposting pesan singkat untuk siapa saja yang mengikuti di *Twitter*, dengan harapan pesan tersebut dapat bermanfaat dan menarik bagi seseorang. Dengan media sosial *Twitter* ini dapat dimanfaatkan untuk menggali banyak informasi salah satunya terkait topik yang hangat dibicarakan saat ini yaitu kebijakan PPKM. Mewabahnya covid-19 di Indonesia telah mengubah banyak tatanan kehidupan masyarakat. Pemerintah berusaha memberlakukan beberapa kebijakan guna menekan penyebaran wabah ini, salah satunya adalah kebijakan PPKM atau Pemberlakuan Pembatasan Kegiatan Masyarakat.

Pemerintah menerapkan PPKM pertama kali pada tanggal 11 Januari-25 Januari 2021 mencakup daerah DKI Jakarta dan 23 kabupaten/kota di enam provinsi pasca libur Natal dan Tahun Baru atau natalu yang memiliki resiko covid-19 yang tinggi akibat tingginya angka kasus covid-19 (Rizal et al., 2021). Walaupun adanya pembatasan sosial ini, pemerintah tetap memberikan ruang bagi kegiatan usaha dan perkantoran untuk beroperasi dengan syarat pembatasan kapasitas dan protokol kesehatan yang ketat harus dilaksanakan. Melalui media sosial *Twitter*, masyarakat dapat membagi ide, mengekspresikan ide, dan mempengaruhi opini publik terkait PPKM. Adanya kebijakan ini memunculkan banyak tanggapan/pandangan pro dan kontra dari masyarakat Indonesia terkait hal tersebut. Beberapa masyarakat mengatakan setuju dan tidak keberatan dengan adanya PPKM walaupun tidak dapat dipungkiri kebijakan tersebut berdampak pada perekonomian masyarakat. Seperti pedagang kecil yang terpaksa tutup lebih awal karena kebijakan yang diberlakukan sehingga penghasilan yang didapat berkurang jauh dari hari biasanya.

Penelitian sebelumnya yang telah mengangkat tema serupa yaitu penelitian Faza Rashif dan lainnya (2020) dengan data cuitan akun bot *Twitter* bertagor #Covid-19 yang diperoleh hasil lima topik teratas, antara lain kondisi dan dampak pandemi saat ini, pertumbuhan penyebaran Covid-19 di Indonesia, himbauan protokol kesehatan menjaga jarak, vaksinasi yang terjadi di beberapa wilayah di Indonesia, dan cara menghadapi covid-19.

Berdasarkan uraian latar belakang yang telah dijelaskan, peneliti akan melakukan kajian tanggapan masyarakat terkait PPKM dengan pemodelan topik metode *Latent Dirichlet Allocation (LDA)* di media sosial *Twitter*. LDA merupakan

salah satu cara untuk mengimplementasikan *topic modelling* dengan model probabilistik generatif di mana setiap dokumen diasumsikan terdiri dari proporsi topik yang berbeda (Habibi et al., 2021). Dengan pengelompokan dokumen ini, dapat dilihat apa saja topik yang dibicarakan masyarakat terkait kebijakan PPKM sehingga dapat menjadi solusi pemerintah dalam menerapkan kebijakan lainnya untuk menekan angka kasus covid-19.

1.2. Rumusan Masalah

Pada penelitian ini, rumusan masalah yang akan diangkat oleh peneliti adalah sebagai berikut :

1. Bagaimana gambaran umum data komentar *Twitter* mengenai perspektif masyarakat terhadap PPKM di Indonesia?
2. Bagaimana hasil analisis sentimen komentar masyarakat pada *Twitter* terkait kebijakan PPKM di Indonesia?
3. Bagaimana hasil *topic modelling* dengan LDA berdasarkan *tweet* perspektif masyarakat tentang PPKM di Indonesia?

1.3. Batasan Masalah

Berdasarkan identifikasi masalah diatas, penulis memberikan batasan masalah dari penelitian sebagai berikut:

1. Data yang digunakan adalah data komentar media sosial *Twitter* dengan Bahasa Indonesia yang diambil dari 1 Januari 2021 sampai 30 November 2021.
2. Data diolah menggunakan *software Microsoft Excel 2013*, *Rstudio* versi 4.10, dan *Phyton (google colab)*.
3. Metode yang digunakan adalah *topic modelling* dengan metode *Latent Dirichlet Allocation (LDA)*

1.4. Tujuan Penelitian

Adapun tujuan dari penelitian yang dilakukan sebagai berikut:

1. Mengetahui gambaran umum data komentar *Twitter* mengenai perspektif masyarakat terhadap PPKM di Indonesia.

2. Mengetahui hasil analisis sentimen komentar masyarakat pada *Twitter* terkait kebijakan PPKM di Indonesia.
3. Mengetahui hasil *topic modelling* dengan LDA berdasarkan *tweet* perspektif masyarakat tentang PPKM di Indonesia.

1.5. Manfaat Penelitian

Adapun manfaat penelitian ini adalah sebagai berikut:

1. Mengetahui topik yang dibahas di *Twitter* mengenai kebijakan PPKM.
2. Berdasarkan hasil penelitian diharapkan dapat membantu pemerintah dalam menindaklanjuti tanggapan masyarakat terkait kebijakan PPKM.

BAB II

TINJAUAN PUSTAKA

Penelitian yang menjadi acuan dalam penelitian kali ini yaitu penelitian terdahulu yang berkaitan dengan PPKM maupun metode *topic modelling* dengan *Latent Dirichlet Allocation* (LDA). Bab ini berfungsi untuk mengembangkan hasil penelitian sebelumnya dan mengetahui hubungan antara penelitian yang dilakukan sekarang dan penelitian terdahulu sehingga dapat bermanfaat dan memiliki arti penting. Adapun tinjauan pustaka yang digunakan adalah sebagai berikut.

Penelitian menggunakan metode LDA dan pemodelan topik: model, aplikasi, survei dimana pada penelitian ini menyebutkan bahwa sebuah teks dapat berupa email, bab buku, posting blog, artikel jurnal, dan segala jenis teks tidak terstruktur. Tema pemodelan dapat memberikan tampilan yang berguna dari koleksi besar dalam hal koleksi sebagai keseluruhan, dokumen individu, dan hubungan antar dokumen. Penelitian ini menyelidiki artikel ilmiah yang sangat tinggi (antara 2003 hingga 2016) terkait dengan *topic modelling* berbasis LDA dalam berbagai ilmu (Jelodar et al., 2019).

Penelitian tentang cuitan masyarakat terhadap covid-19 menggunakan metode LDA didapatkan hasil topik teratas sebanyak 5 topik antara lain membahas tentang dampak dan kondisi pandemi saat ini, himbauan untuk menjaga kesehatan dengan jaga jarak, perkembangan penyebaran covid-19 di Indonesia, vaksinasi covid-19 di Indonesia, dan cara untuk menghadapi covid-19 (Rashif et al., 2021).

Penelitian analisis tren topik dengan LDA yang bertujuan untuk menemukan topik utama keluhan warga Denpasar dengan cepat dan efisien. Didapatkan hasil trend suatu topik berada pada topik 4 didasarkan pada yang tertinggi probabilitas topik yaitu sebesar 0,055 yang dapat dimaknai banyak warga Denpasar yang mengeluh jalan rusak dan diminta untuk memperbaiki jalan (Destarani et al., 2019).

Penelitian judul “Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial *Twitter* Menggunakan *Naive Bayes Clasifiers*” yang bertujuan untuk mengetahui bagaimana sentimen masyarakat terhadap penerapan kebijakan PPKM di Indonesia. Didapatkan hasil kesimpulan opini masyarakat mengenai kebijakan PPKM menghasilkan klasifikasi polaritas

positif sebesar 99% dan polaritas negatif sebesar 1%. Kemudian dihasilkan bahwa kata “darurat” adalah kata dengan frekuensi tertinggi penggunaannya pada pengguna *Twitter* (Taofik Krisdiyanto, 2021).

Penelitian deteksi topik dengan metode LDA tentang tokoh publik politik menggunakan data tweet sosial media *Twitter* yang akan berfungsi sebagai pemodelan topik tweet bengkok, dan TF-IDF untuk mengetahui tweet mana yang berisi kata-kata dalam LDA yang akan menjadi topik referensi. Ini akan menghasilkan deteksi yang relevan topik berdasarkan tweet tentang tokoh politik publik (Hikmah et al., 2020).

Penelitian yang dilakukan oleh Risa Wati dan Siti Ernawati menggunakan analisis sentimen dengan algoritma SVM pada data komentar masyarakat tentang PPKM. Penelitian tersebut menunjukkan bahwa algoritma SVM dengan kernel linier terbukti menjadi algoritma pengklasifikasi text yang baik pada analisis sentimen pemberlakuan PPKM wilayah Jawa dan Bali dan analisis sentimen penerapan PPKM untuk wilayah Jawa dan Bali terbukti berhasil menekan angka penyebaran Covid-19 (Wati et al., 2021).

Penelitian dengan analisis SWOT tentang PSBB dan PPKM terhadap dampak ekonomi diharapkan mampu menekan penyebaran virus covid-19 yang semakin mengkhawatirkan. Dalam analisis SWOT ini dibahas tentang PPKM yang tentunya memiliki kelebihan dan kekurangan sehingga akan memberikan gambaran keberhasilan kebijakan yang diambil oleh pemerintah (Ilyas, 2021).

Penelitian melalui Instagram, komik bertema PPKM tersebar dan mendapat banyak perhatian dari masyarakat. Penelitian ini bertujuan untuk mengetahui bagaimana penggunaan komik sebagai kritik sosial dalam pelaksanaan PPKM dengan metode *paradigma critical metaphor analysis*. Hasil penelitian menunjukkan 50 komik mengkritik kebijakan PPKM dari 20 akun. Para pencipta mengangkat enam metafora signifikan, antara lain PPKM sebagai berbagai jenis model, PPKM sebagai bentuk keserakahan, PPKM sebagai penjara, Covid-19 digambarkan dalam kelas sosial, Pemerintah adalah pahlawan PPKM sebagai bentuk arogansi (Audinovic & Nugroho, 2021).

Penelitian lain dengan metode analisis deskriptif dan kuantitatif terkait tingkat pendidikan dan kemiskinan yang terjadi di Aceh didapatkan hasil bahwa

indikator tingkat pendidikan penduduk Aceh sudah jauh lebih baik dan bahkan berada di atas level nasional dibandingkan dengan tingkat pendidikan penduduk di 34 Provinsi lainnya di Indonesia. Namun mutunya masih sangat memprihatinkan (Majid, 2014).

Penelitian dengan metode *descriptive statistics* dan *sentiment analysis* pada studi rekayasa perangkat lunak dengan melakukan studi sederhana tentang perbedaan waktu penyelesaian masalah untuk teks positif, negatif dan netral. Pada penelitian ini mempelajari dampak dari pilihan alat analisis sentimen ketika melakukan studi rekayasa perangkat lunak. Telah diamati bahwa tidak hanya alat yang dianggap tidak sesuai dengan pelabelan manual, tetapi juga mereka tidak setuju satu sama lain, bahwa ketidaksepakatan ini dapat menyebabkan kesimpulan yang berbeda dan hasil yang dipublikasikan sebelumnya tidak dapat direplikasi ketika alat analisis sentimen yang berbeda digunakan (Jongeling et al., 2017).

Tabel 2.1 Deskripsi Penelitian Terdahulu

| No | Nama peneliti | Judul Penelitian | Metode | Persamaan | Perbedaan |
|-----------|---|---|--|---|---|
| 1. | Hamed Jelodar, Yongli Wang, Chi Yuan, Xia Feng, Xiahui Jiang, Yanchao Li, Liang Zhao (2018) | <i>Latent Dirichlet Allocation (LDA) and Topic modeling: models, applications, a survey</i> | <i>Latent Dirichlet Allocation (LDA)</i> | Menggunakan metode yang sama | Data yang digunakan mengenai PPKM pada media sosial <i>Twitter</i> |
| 2. | Faza Rashif, Goldio Ihza Perwira Nirvana, Muhammad Alif Noor, Nur Aini Rakhmawati (2021) | Implementasi LDA untuk Pengelompokan Topik Cuitan Akun Bot <i>Twitter</i> bertagar #Covid-19 | <i>Latent Dirichlet Allocation (LDA)</i> | Membahas studi kasus mengenai covid-19 pada <i>Twitter</i> dengan LDA | Data yang digunakan mengenai PPKM pada media sosial <i>Twitter</i> |
| 3. | Aulia Rizki Destarani, Isnandar Slamet, Sri Subanti (2019) | <i>Trend Topic Analysis using Latent Dirichlet Allocation (LDA)</i> | <i>Latent Dirichlet Allocation (LDA)</i> | Menggunakan metode yang sama | Data yang digunakan mengenai PPKM pada media sosial <i>Twitter</i> |
| 4. | Faizun Nuril Hikmah, Setio Basuki, Yufis Azhar (2020) | Deteksi Topik Tentang Tokoh Publik Politik Menggunakan <i>Latent Dirichlet Allocation</i> | <i>Latent Dirichlet Allocation (LDA)</i> | Menggunakan metode yang sama | Data yang digunakan mengenai PPKM pada media sosial <i>Twitter</i> |
| 5. | Risa Wati, Siti Ernawati (2021) | Analisis Sentimen Persepsi Publik Mengenai PPKM Pada <i>Twitter</i> Berbasis SVM Menggunakan Python | Algoritma SVM | Membahas studi kasus mengenai PPKM pada <i>Twitter</i> | Metode yang digunakan adalah LDA pada media sosial <i>Twitter</i> . |

| No | Nama peneliti | Judul Penelitian | Metode | Persamaan | Perbedaan |
|----|---|---|---|--|---|
| 6. | Taofik Krisdiyanto, Erry Maricha Oki Nurharyanto (2021) | Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial <i>Twitter</i> Menggunakan <i>Naive Bayes Clasifiers</i> | <i>Naive Bayes Clasifiers</i> | Membahas studi kasus mengenai PPKM pada <i>Twitter</i> | Metode yang digunakan adalah LDA pada media sosial <i>Twitter</i> . |
| 7. | Fadli Ilyas (2021) | Analisis SWOT Kebijakan Pembatasan Sosial Berskala Besar (PSBB) dan Pemberlakuan Pembatasan Kebijakan Masyarakat (PPKM) Terhadap Dampak Ekonomi Di Tengah Upaya Menekan Laju Pandemi Covid-19 | Analisis SWOT | Membahas studi kasus mengenai PPKM | Metode yang digunakan adalah LDA pada media sosial <i>Twitter</i> |
| 8. | Vizcardine Audinovic, Rio Satria Nugroho (2021) | <i>ComicasSocial Media Criticism of Community Activities Restrictions Enforcement Due to COVID-19 in Indonesia</i> | <i>Paradigma critical metaphor analysis</i> | Membahas studi kasus mengenai PPKM | Metode yang digunakan adalah LDA pada media sosial <i>Twitter</i> |
| 9. | M. Shabri Abd. Majid (2014) | Analisis Tingkat Pendidikan dan Kemiskinan di Aceh | Analisis deskriptif dan kuantitatif | Membahas metode analisis deskriptif | Metode yang digunakan adalah analisis deskriptif dan LDA pada media sosial <i>Twitter</i> |

| No | Nama peneliti | Judul Penelitian | Metode | Persamaan | Perbedaan |
|-----------|--|--|---|--|---|
| 10. | Robbert Jongeling, Proshanta Sarkar, Subhajit Datta, Alexander Serebrenik | <i>On Negative Results When Using Sentiment Analysis Tools for Software Engineering Research</i> | <i>Descriptive statisics dan sentiment analysis</i> | Membahas metode analisis deskriptif | Metode yang digunakan adalah analisis deskriptif dan LDA pada media sosial <i>Twitter</i> |

BAB III

LANDASAN TEORI

3.1. Kebijakan PPKM

Pemerintah mengeluarkan surat edaran terkait instruksi kebijakan pemberlakuan pembatasan kegiatan masyarakat (PPKM) pertama kali yang berlaku pada tanggal 11 Januari - 25 Januari 2021. Daerah yang menjadi kebijakan tersebut merupakan daerah yang memiliki risiko covid-19 yang tinggi mencakup DKI Jakarta dan 23 kabupaten/kota di enam provinsi. Sebelum adanya kebijakan PPKM, pemerintah memperkenalkan sebutan lain pada awal pandemi yaitu PSBB dan PPKM Mikro yang mulai diberlakukan pada 17 April 2020. Kemudian pemerintah memberlakukan istilah baru untuk menekan penyebaran covid-19 pada daerah beresiko tinggi dengan nama PPKM (Rizal et al., 2021).

Sejak awal tahun 2021, kebijakan pemerintah Indonesia menerapkan pembatasan kegiatan masyarakat (PPKM) dalam rangka menghadapi pandemi covid-19 di Indonesia. Berikut **Tabel 3.1** jenis tahapan PPKM beserta waktu dan wilayah penerapannya.

Tabel 3.1 Tahapan Kebijakan PPKM

| Jenis | Tanggal | Wilayah |
|---------------|------------------------------|---|
| PPKM jilid I | 11 Januari - 25 Januari 2021 | Terdiri dari 7 provinsi (DKI Jakarta, Jawa Barat, Banten, Jawa Tengah, Yogyakarta, Jawa Timur, Bali) |
| PPKM jilid II | 26 Januari - 8 Februari 2021 | Pulau Jawa dan Bali yaitu DKI Jakarta, Banten, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, dan Bali. |
| PPKM mikro | 22 Juni - 5 Juli 2021 | 122 kabupaten dan kotamadya yang ada di Pulau Jawa dan Bali terdiri dari 48 kabupaten dan kota dengan penilaian level 4 dan 74 kabupaten dan kota dengan penilaian level 3. |

| Jenis | Tanggal | Wilayah |
|--------------|----------------|--|
| PPKM darurat | 3-25 Juli 2021 | Jawa dan Bali, serta 15 kabupaten/kota luar Jawa-Bali yang wajib melaksanakan PPKM Darurat yaitu Kota Tanjungpinang, Kota Singkawang, Kota Padang Panjang, Kota Balikpapan, Kota Bandar Lampung, Kota Pontianak, Kabupaten Manokwari, Kota Sorong, Kota Batam, Kota Bontang, Bukittinggi Kota, Kabupaten Berau, Kota Padang, Kota Mataram, dan Kota Medan. |

Pelaksanaan PPKM terbagi menjadi beberapa tahapan, antara lain jilid I dan jilid II, PPKM mikro, dan PPKM darurat. PPKM jilid diberlakukan sesuai dengan Imendagri Nomor 1 Tahun 2021 tentang Permohonan Pembatasan Pengendalian Penyebaran Corona Virus Disease 2019 (covid-19) dan Wilayah terdiri dari 7 provinsi (DKI Jakarta, Jawa Barat, Banten, Jawa Tengah, Yogyakarta, Jawa Timur, Bali). PPKM diadakan sebanyak dua jilid dan ternyata hasilnya tidak efektif, PPKM diubah menjadi PPKM berbasis mikro.

PPKM berbasis mikro dimulai pada 22 Juni 2021 dan berakhir pada 5 Juli 2021, sesuai Permendagri Inpres Dalam Negeri No. 2-14.17 dan 23 Tahun 2021. Sebagaimana telah disebutkan sebelumnya, PPKM Mikro adalah berlaku di sejumlah tempat di tujuh provinsi. Namun berbeda dengan PPKM, PPKM mikro meliputi pengaturan pendirian posko penanganan covid-19 di tingkat desa dan desa. Pusat perbelanjaan / mal beroperasi dengan jadwal yang lebih fleksibel, yang sampai dengan pukul 21.00 WIB, dan dengan jadwal yang lebih fleksibel yaitu 50 persen bekerja dari kantor dan 50 persen bekerja dari rumah. PPKM Mikro diperpanjang berkali-kali oleh pemerintah setelah ditahan selama dua minggu (Miharja et al., 2021).

PPKM darurat sesuai dengan instruksi Mendagri berlaku mulai 3 hingga 20 Juli 2021 kemudian diperpanjang hingga 25 Juli sebagai PPKM level 4 (M. Yoserizal Saragih, 2021). Hal ini bertujuan untuk mengurangi kasus harian seperti disini konfirmasi kasus menjadi kurang dari 10.000 per hari. PPKM Darurat covid-19 di kabupaten dan kota di wilayah Jawa dan Bali dengan kriteria level 3 dan level 4 sebagaimana dimaksud dalam diktum pertama melaksanakan kegiatan seperti pelaksanaan belajar-mengajar kegiatan dilakukan secara *online*, pelaksanaan kegiatan di sektor non esensial 100% bekerja dari rumah (WFH), pelaksanaan kegiatan di sektor esensial seperti perbankan 50% maksimal *staff work from office* (WFO) dengan protokol kesehatan ketat, tempat ibadah ditutup sementara (Miharja et al., 2021).

3.2. Media Sosial *Twitter*

Media sosial adalah teknologi berbasis komputer yang memfasilitasi berbagai ide, pemikiran, dan informasi melalui pembangunan jaringan dan komunitas virtual. Media sosial digunakan sebagai cara untuk berinteraksi dengan teman, keluarga, ataupun orang lain tetapi kemudian diadopsi oleh bisnis untuk menjangkau pelanggan dengan memanfaatkan metode komunikasi baru. Kekuatan media sosial adalah kemampuan untuk terhubung dan berbagi informasi dengan siapa saja secara bersamaan. Seiring berkembangnya teknologi dan komunikasi menjadikan media sosial hadir sebagai media baru untuk memudahkan manusia dalam berkomunikasi. (Bin Muhammad Alkatiri et al., 2020).

Salah satu media sosial yang sering digunakan untuk mempertukarkan wacana atau pemikiran adalah *Twitter*. *Twitter* merupakan situs web yang dimiliki dan dioperasikan oleh *Twitter, Inc.*, yang menawarkan layanan jejaring sosial dan *microblogging*. Disebut *microblogging* karena situs ini memungkinkan penggunanya untuk mengirim dan membaca pesan pengguna lain seperti pada umumnya tetapi dibatasi hanya 140 karakter yang ditampilkan di halaman beranda pengguna. *Twitter* memiliki karakteristik dan format penulisan yang unik dengan simbol atau aturan khusus. (Novantirani et al., 2015).

Pada *Twitter*, aktivitas utama yang dilakukan adalah memposting pesan atau sesuatu yang singkat (*tweet*) melalui web atau seluler. Panjang maksimum *tweet*

adalah 280 karakter. *Twitter* adalah sumber yang hampir tidak terbatas yang digunakan dalam klasifikasi teks. Ada banyak karakteristik dari tweet *Twitter*. Pesan di *Twitter* memiliki banyak atribut unik, yang membedakannya dari media sosial lainnya:

1. *Twitter* memiliki panjang karakter maksimal 280 karakter.
2. *Twitter* menyediakan data yang dapat diakses secara bebas menggunakan *Twitter* API, sehingga memudahkan untuk mengumpulkan *tweet* dalam jumlah banyak.
3. Model bahasa - Pengguna *Twitter* memposting pesan melalui banyak media yang berbeda. Frekuensi ejaan, bahasa gaul, dan singkatan lebih tinggi dibandingkan media sosial lainnya.
4. Pengguna *Twitter* mengirim *tweet* tentang berbagai topik yang disesuaikan dengan topik tertentu dan berlaku secara umum (Novantirani et al., 2015).

3.3. *Twitter Web Scrapping*

Pereira et al. menyatakan bahwa *web scraping* adalah proses mengekstrak data dari internet dengan metode atau teknik apapun. Pengikisan web juga membantu otomatisasi web dalam banyak hal, termasuk data cuaca, deteksi perubahan web, dan perbandingan harga situs web *online*. *Web scraping* dapat mengubah data tidak terstruktur menjadi data terstruktur yang kemudian disimpan dan divalidasi ke dalam *database*. Pada intinya *web scraping* adalah mengumpulkan data, menyimpan data, dan melakukan validasi data. Langkah penting berikutnya adalah analisis data, sehingga data yang telah dilakukan validasi dapat diinterpretasikan menjadi informasi yang lebih baik.

Twitter developers API bekerja sangat mirip dengan API Pengembang *Facebook*. Kita harus membuat akun *Twitter* dan kemudian pergi ke situs resmi Pengembang *Twitter* (<https://developer.Twitter.com/en.html>) untuk mendaftarkan aplikasi kita dan kemudian mendapatkan ID Aplikasi dan Token Akses (Dewi et al., 2019).

3.4. *Machine Learning*

Machine learning adalah jenis kecerdasan buatan yang berfokus pada pengembangan computer program yang bisa berubah ketika terkena data baru. Ini

menggunakan model dan informasi komputer diperoleh dari data masa lalu dan sebelumnya untuk membantu proses klasifikasi, prediksi dan deteksi. Pada tingkat paling dasar, pembelajaran mesin mengacu pada semua jenis program komputer yang dapat "belajar" dengan sendirinya tanpa harus diprogram secara eksplisit oleh manusia (Wehle, 2017).

Machine Learning (ML) merupakan cabang algoritma dan pendekatan dalam AI yang banyak digunakan untuk meniru kecerdasan manusia dengan belajar dari lingkungan sekitarnya. ML mencoba belajar dan menggeneralisasi dengan cara menirukan bagaimana proses manusia atau makhluk cerdas. Dalam ML terdapat dua aplikasi utama yaitu, klasifikasi dan prediksi. Karakteristik dari ML adalah adanya proses pelatihan, pembelajaran, atau *training*. Oleh karena itu, ML membutuhkan data untuk dipelajari yang disebut sebagai data *training*. Klasifikasi adalah metode dalam ML yang digunakan oleh mesin untuk memilah atau mengelompokkan objek berdasarkan ciri tertentu seperti bagaimana manusia mencoba membedakan benda satu dengan yang lain (Ahmad Hania, 2017).

Berdasarkan penelitian Shagan Sah tahun 2020 terdapat 2 kategori utama dari algoritma *machine learning*, yaitu :

1. *Supervised Learning*

Supervised learning diterapkan ketika data dalam bentuk variabel *input* dan nilai target *output*. Algoritma ini mempelajari fungsi pemetaan dari input ke output. Ketersediaan sampel data berlabel skala besar membuatnya pendekatan yang mahal untuk tugas-tugas di mana data langka. Ini pendekatan secara luas dapat dibagi menjadi dua kategori utama yaitu klasifikasi dan regresi (Sah, 2020).

2. *Unsupervised Learning*

Unsupervised learning diterapkan ketika data tersedia hanya dalam bentuk *input* dan tidak ada variabel *output* yang sesuai. Algoritma semacam itu memodelkan yang mendasarinya pola dalam data untuk mempelajari lebih lanjut tentang karakteristiknya. Salah satu jenis utama dari algoritma *unsupervised learning* adalah pengelompokan. Dalam teknik ini, kelompok yang melekat dalam data adalah ditemukan dan kemudian digunakan untuk

memprediksi *output* untuk *input* yang tidak terlihat. Contoh dari teknik ini adalah untuk memprediksi pelanggan perilaku pembelian (Sah, 2020).

3.5. Data Mining

Perkembangan teknologi informasi telah menghasilkan sejumlah besar *database* dan data yang sangat besar di berbagai bidang. Penelitian dalam basis data dan teknologi informasi telah memunculkan pendekatan untuk menyimpan dan memanipulasi data berharga ini untuk pengambilan keputusan lebih lanjut. *Data mining* adalah proses yang digunakan untuk menemukan pola yang berguna dari sejumlah besar data. Disebut juga sebagai proses penemuan pengetahuan, penggalian pengetahuan dari data, ekstraksi pengetahuan atau analisis data/pola. Setelah pola-pola ini ditemukan, selanjutnya dapat digunakan untuk membuat keputusan tertentu untuk pengembangan bisnis. Tiga langkah yang terlibat adalah

1. *Exploration* : Pada langkah pertama eksplorasi data, data dibersihkan dan diubah menjadi bentuk lain, dan variabel penting dan kemudian sifat data berdasarkan masalah ditentukan.
2. *Pattern Identification* : Setelah data dieksplorasi, disempurnakan, dan ditentukan untuk variabel spesifik, langkah kedua adalah membentuk identifikasi pola. Identifikasi dan pilih pola yang membuat prediksi terbaik.
3. *Deployment*: Pola dikerahkan untuk hasil yang diinginkan (Dey et al., 2017).

3.6. Text Mining

Text mining adalah proses mendapatkan informasi berkualitas tinggi dari teks. *Text mining* biasanya melibatkan proses penataan teks *input*, menemukan pola di dalam teks terstruktur data, dan akhirnya evaluasi dan interpretasi *output*. Tugas *text mining* yang umum termasuk teks kategorisasi, pengelompokan teks, peringkasan dokumen, ekstraksi kata kunci, dan lain-lain.

Sebagai salah satu bidang utama dalam data mining, *text mining* bertujuan untuk menemukan pengetahuan yang sebelumnya tidak diketahui tetapi berpotensi berguna dari data teks tidak terstruktur atau semi terstruktur. *Text mining* juga menghadapi masalah seperti data dalam jumlah besar, dimensi tinggi, data dan struktur yang terus berubah, dan "noise" data. Tidak seperti *data mining*, yang

terutama memproses data terstruktur, data yang digunakan oleh *text mining* umumnya tidak terstruktur, atau setidaknya teks semi terstruktur (Fitri et al., 2019).

Pada proses *text mining* perlu dilakukan beberapa tahapan untuk mengolah sumber data baik yang terstruktur, terstruktur sebagian dan yang tidak terstruktur dari beberapa sumber. Data-data tersebut perlu dilakukan proses awal atau disebut sebagai *preprocessing text* yang bertujuan mengolah data awal yang masih bermacam-macam atau tidak teratur untuk dijadikan sebuah data teratur yang dapat diterapkan beberapa metode *text mining* (H., 2015). *Preprocessing* merupakan tahapan yang sangat penting dalam teknik dan aplikasi *text mining*. Ini adalah langkah pertama dalam proses *text mining*. Terdapat beberapa langkah-langkah *preprocessing* yaitu, *cleaning*, *case folding*, *tokenizing*, *stop-word removal* dan *stemming*. Pada penelitian ini tahap *stop-word removal* dan *stemming* menggunakan pustaka sastrawi.

a. *Cleaning*

Pembersihan data adalah teknik penting untuk prapemrosesan data. Ini adalah proses teknik *data mining*. Ini menghapus data kesalahan buruk dan mengurangi informasi data yang tidak perlu. Data yang hilang juga termasuk dalam teknik pembersihan data (T. Nikil Prakash, 2019). Langkah pembersihan akan menghilangkan kata atau karakter termasuk HTML, emoticon, hashtag (#), username (@namapengguna), dan url (Siti Qomariyaha, 2019).

b. *Case Folding*

Pada sebuah dokumen, penggunaan huruf kapital atau sejenisnya terkadang tidak memiliki kesamaan, hal ini dapat terjadi karena kesalahan penulisan. Pada proses *case folding* ini bertujuan untuk mengubah semua huruf pada a dokumen teks menjadi huruf kecil, misalnya kata "Wifi" menjadi "wifi" (Rosid et al., 2020).

c. *Tokenizing*

Sebuah dokumen teks terdiri dari sekumpulan kalimat, proses *tokenization* memecah dokumen menjadi beberapa bagian dari kata-kata yang disebut token (Rosid et al., 2020).

d. *Filtering*

Pada tahap *filtering* digunakan algoritma *stop-word removal*. Penghapusan kata bertujuan untuk menghilangkan kata-kata yang tidak penting, proses ini dapat mengurangi dimensi ruang yang terlihat berat. Kata-kata umum dalam dokumen teks seperti pakaian, preposisi dan kata benda dll, yang tidak memberi arti pada dokumen. Contoh kata "yang", "dan", "Anda", "sampai" dihapus dari dokumen karena tidak diukur sebagai kata kunci dalam aplikasi penambangan teks. Dalam penelitian ini menggunakan perpustakaan sastrawi. Gambar 3 adalah stop-word cuplikan kode sumber (Rosid et al., 2020).

e. *Stemming*

Stemming adalah proses penguraian bentuk suatu kata ke dalam bentuk kata dasarnya. Lebih mudah dikatakan, proses mengubah kata yang memiliki imbuhan menjadi kata dasar. Proses *stemming* ini adalah proses terpenting dalam tahap penambangan teks. Hasil *stemming* yang baik dapat mempengaruhi baik atau tidaknya aplikasi penambangan teks. Gambar 4 adalah potongan source code menggunakan library sastrawi. perpustakaan Sastrawi untuk bahasa pemrograman PHP dapat di install dengan langkah-langkah sebagai berikut :

- 1) Buka terminal (baris perintah) dan navigasikan ke direktori proyek Anda
- 2) *Download Composer* agar *file composer.phar* berada di direktori tersebut.
- 3) Tambahkan sastrawi ke *file composer.json* (Rosid et al., 2020).

3.7. Analisis Sentimen

Analisis sentimen merupakan proses yang mengotomatisasi penggalian sikap, pandangan, pendapat, dan emosi dari teks, pidato, *tweet*, dan sumber basis data melalui *Natural Language Processing* (NLP). Analisis sentimen melibatkan pengklasifikasian pendapat dalam teks ke dalam kategori seperti "positif" atau "negatif" atau "netral". Ini juga disebut sebagai analisis subjektivitas, opini

penambahan, dan ekstraksi penilaian. Kata-kata opini, sentimen, pandangan dan keyakinan digunakan bergantian tetapi ada perbedaan di antara mereka.

- *Opinion* : suatu kesimpulan terbuka untuk diperdebatkan (karena para ahli berbeda pendapat)
- *View* : opini subjektif
- *Belief* : penerimaan yang disengaja dan persetujuan intelektual
- *Sentiment* : opini yang mewakili perasaan seseorang

Contoh terminologi untuk Analisis Sentimen adalah sebagai berikut,

<KALIMAT> = Cerita filmnya lemah dan membosankan

<OPINI PEMEGANG> =<penulis>

<OBJEK> = <film>

<FITUR> = <cerita>

<OPINI>= <lemah><membosankan>

<POLARITAS> = <negatif>

(A. & Sonawane, 2016)

3.8. Term Frequency - Inverse Document Frequency (TF-IDF)

TF-IDF telah banyak digunakan dalam bidang pencarian informasi dan penambahan teks untuk mengevaluasi hubungan untuk setiap kata dalam kumpulan dokumen. Secara khusus digunakan untuk mengekstrak kata-kata inti (yaitu, kata kunci) dari dokumen, menghitung derajat yang sama di antara dokumen, menentukan peringkat pencarian, dan sebagainya. TF dalam TF-IDF berarti kemunculan kata-kata tertentu dalam dokumen. Kata-kata dengan nilai TF yang tinggi memiliki arti penting dalam dokumen. Di sisi lain, DF menyiratkan berapa kali kata tertentu muncul dalam kumpulan dokumen. disini kemudian dihitung kemunculan kata dalam banyak dokumen, tidak hanya dalam satu dokumen. Kata-kata dengan nilai DF yang tinggi tidak terlalu penting karena biasanya muncul di semua dokumen. Oleh karena itu, IDF yang merupakan kebalikan dari DF digunakan untuk mengukur pentingnya kata-kata dalam semua dokumen. Nilai IDF yang tinggi berarti kata-kata langka di semua dokumen, mengakibatkan peningkatan kepentingan (Gil, 2019).

Menurut Siregar dkk (2017) terdapat empat cara penghitungan untuk mencari nilai TF :

1. *Raw TF*

Raw TF menentukan bobot suatu dokumen terhadap istilah dengan menghitung jumlah kemunculan suatu kata tersebut pada suatu dokumen. Contohnya ketika nilai TF disini adalah 1 maka dalam suatu dokumen muncul suatu kata sebanyak 1 kali.

2. *Logarithmic TF*

Logarithmic TF melakukan log terhadap TF untuk mengurangi tingkat kepentingan kemunculan kata dalam menghitung bobot dokumen terhadap suatu istilah dengan persamaan berikut :

$$TF = 1 + \log(TF) \quad (3. 1)$$

3. *Binnary TF*

Binary TF memberi nilai 0 dan 1 untuk menyamakan bobot dokumen terhadap istilah. Nilai 1 menyatakan suatu istilah muncul minimal satu kali dalam suatu dokumen, sementara 0 menyatakan tidak adanya istilah yang muncul dalam suatu dokumen.

4. *Augmented TF*

Augmented TF menyamakan bobot dokumen terhadap istilah dengan memberikan rentang antara 0.5 hingga 1 sebagai bobot dokumen. *Augmented TF* dapat dihitung dengan persamaan:

$$TF = 0,5 + 0,5 \times \frac{TF}{\max(TF)} \quad (3. 2)$$

Selajutnya IDF merupakan frekuensi dokumen terbalik dari kata di satu set dokumen. hal ini menunjukkan seberapa umum atau jarang sebuah kata di seluruh kumpulan dokumen. Semakin dekat ke 0, semakin umum sebuah kata. IDF dihitung dengan mengambil jumlah total dokumen, lalu membaginya dengan jumlah dokumen yang berisi kata, dan menghitung logaritma.

IDF dihitung dengan persamaan :

$$IDF_j = \ln\left(\frac{D}{df_j}\right) \quad (3. 3)$$

Keterangan :

IDF_j : *Invers document frequency* dari term j

D : Total dokumen

df_j : Total dokumen dari term j

Sehingga didapatkan penghitungan TF-IDF sebagai berikut :

$$W_{ij} = TF_{ij} \times IDF_j \quad (3.4)$$

$$W_{ij} = TF_{ij} \times \ln\left(\frac{D}{df_j}\right) \quad (3.5)$$

Keterangan :

W_{ij} : Bobot dari term i terhadap dokumen j

TF_{ij} : Frekuensi munculnya *term i* di dalam dokumen j

IDF_j : *Invers document frequency* dari *term j*

D : Total dokumen

df_j : Total dokumen dari *term j*

(Nisrina, 2020)

Jika frekuensi kemunculan *term* sama dengan jumlah dokumen, maka hasil perhitungan $IDF_j = 0$. Untuk menghindari hasil $w_{ij} = 0$, maka pada hasil perhitungan IDF akan ditambahkan dengan nilai 1 (satu) seperti berikut :

$$W_{ij} = TF_{ij} \times \left(\ln\left(\frac{D}{df_j}\right) + 1\right) \quad (3.6)$$

(Apriani et al., 2021)

3.9. *Natural Language Processing (NLP)*

NLP adalah area penelitian dan aplikasi yang melakukan eksplorasi tentang bagaimana komputer dapat digunakan untuk memahami dan memanipulasi teks bahasa alami. NLP bertujuan untuk mengumpulkan pengetahuan tentang bagaimana manusia memahami dan menggunakan bahasa sehingga alat dan teknik yang sesuai dapat dikembangkan untuk membuat sistem komputer memahami dan memanipulasi bahasa alami untuk melakukan tugas yang disukai. Dasar-dasar NLP terletak pada sejumlah disiplin ilmu, yaitu. ilmu komputer dan informasi, linguistik, matematika, teknik listrik dan elektronik, psikologi, dll. Aplikasi NLP mencakup

sejumlah bidang studi, seperti terjemahan mesin, pemrosesan dan peringkasan teks bahasa alami, antarmuka pengguna, multibahasa dan pencarian informasi lintas bahasa (CLIR), pengenalan suara, kecerdasan buatan dan sistem pakar dan sebagainya (Kannan, S., Gurusamy, V., Vijayarani, S., Ilamathi, J. & Nithya, 2015).

3.8.1 Topic Modelling

Topic modelling adalah algoritma penting yang digunakan dalam *text mining*. *Topic modelling* adalah model probabilistik yang menemukan tema utama dalam kumpulan dokumen. Ide dasarnya adalah memperlakukan dokumen sebagai campuran topik dalam topik model, dan setiap topik dipandang sebagai distribusi probabilitas dari kata-kata. Saat menggunakan model topik sebagai alat penambahan teks, setiap topik dipandang sebagai kumpulan kata-kata, dan masing-masing dokumen dapat dilihat sebagai satu set topik dengan yang berbeda proporsi tergantung pada frekuensi kemunculan istilah (Yang & Zhang, 2018).

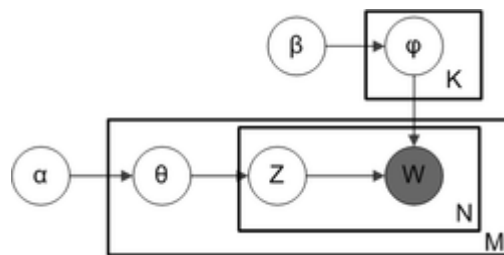
Metode *topic modelling* umumnya digunakan untuk mengatur, memahami, mencari, dan meringkas arsip elektronik berukuran besar secara otomatis. *Topic Modelling* adalah teknik *text mining* untuk menemukan 'subjek' abstrak dari serangkaian dokumen. Sebuah dokumen umumnya ditulis pada satu topik, dan dengan demikian, kata-kata yang terkait dengan topik akan muncul lebih sering daripada kata-kata lain dalam dokumen. Misalnya, dalam dokumen di subjek anjing, kata "anjing" dan "tulang" akan lebih sering muncul, sementara diasumsikan bahwa dokumen tentang kucing akan lebih sering berisi kata-kata "kucing" dan "meong". Model topik, berbicara kasar, mengikat kata "anjing" dan "tulang" di bawah satu topik, dan "kucing" dan "meong" di bawah topik lain (Jelodar et al., 2019).

3.8.2 Latent Dirichlet Allocation (LDA)

LDA adalah probabilistik generatif model topik di mana setiap dokumen direpresentasikan sebagai campuran acak dari topik laten dan setiap topik direpresentasikan sebagai distribusi di atas set tetap kata. LDA bertujuan untuk mengidentifikasi struktur topik laten yang mendasari berdasarkan data yang diamati. Di LDA, kata-kata dari setiap dokumen adalah data yang diamati. Untuk

setiap dokumen dalam korpus, kata-kata dihasilkan. Di LDA, kata adalah data diskrit dari kosakata yang diindeks oleh $\{1, \dots, V\}$, urutan N kata $w = (w_1, w_2, \dots, w_n)$ dan corpus adalah kumpulan dari M dokumen dilambangkan dengan $D = (w_1, w_2, \dots, w_M)$ (Habibi et al., 2021).

Proses LDA dapat dimodelkan dengan model grafis Bayesian tiga tingkat, di mana variabel acak diwakili oleh node dan kemungkinan ketergantungan antara variabel diwakili oleh tepi, seperti yang digambarkan pada **Gambar 3.1** Dalam representasi ini, mengacu pada parameter *dirichlet*, mengacu pada variabel topik tingkat dokumen, z mengacu pada tugas topik per kata, w mengacu pada kata yang diamati dan mengacu pada topik. Seperti dapat diamati dari representasi tiga lapis, parameter dan diambil sampelnya satu kali sambil menghasilkan korpus, topik tingkat dokumen variabel sampel untuk setiap dokumen dan variabel tingkat kata adalah sampel untuk setiap kata dari dokumen.



Gambar 3.1 Representasi grafis dari LDA
(Kandukuri & Haragopal, 2020)

Keterangan :

K = jumlah topik

M = banyaknya dokumen

N = banyaknya kata dalam dokumen tertentu

α = probabilitas pada distribusi topik per-dokumen

β = probabilitas pada distribusi kata per topik

θ = distribusi untuk dokumen d

ϕ = distribusi kata untuk topik k

Z = topik untuk kata yg ke- n dari suatu dokumen

W = kata khusus

Gambar 3.1 menunjukkan tiga tingkat representasi LDA. Tingkat pertama adalah parameter tingkat korpus yang diwakili oleh simbol α dan β . Tingkat korpus ini parameter diasumsikan diambil sampelnya sekali dalam proses menghasilkan korpus. Kedua, variabel tingkat dokumen (θ), sampel sekali dari setiap dokumen. Akhirnya, tingkat kata variabel yang disimbolkan dengan z dan w . Tingkat kata variabel diambil sampelnya sekali untuk setiap kata di setiap dokumen (Muhammad Habibi, 2020).

Peluang LDA dihitung dengan persamaan :

$$p(w, z, \theta, \varphi | \alpha, \beta) = p(\varphi | \beta) p(\theta | \alpha) p(z | \theta) p(w | \varphi_k) \quad (3. 7)$$

Implementasi tidak dapat menerapkan LDA karena variabel z tidak diketahui. Lalu sulitnya untuk menemukan hubungan antara z dan w karena kata tersebut mungkin mengandung lebih dari satu elemen. Juga menghasilkan $p(z | w)$:

$$p(\vec{z} | \vec{w}) = \frac{p(\vec{z} | \vec{w})}{p(\vec{w})} = \frac{\prod_{i=1}^W p(z_i, w_i)}{\prod_{i=1}^W \sum_{k=1}^K p(z_i=k, w_i)} \quad (3. 8)$$

(Rashif et al., 2021)

LDA adalah alat yang dibedakan untuk distribusi topik laten dengan korpus besar. Oleh karena itu memiliki kemampuan untuk mengidentifikasi sub-topik untuk area teknologi yang terdiri dari banyak paten, dan mewakili setiap paten dalam berbagai distribusi topik. Dengan LDA, istilah dalam kumpulan dokumen, menghasilkan kosakata yang kemudian diterapkan untuk menemukan topik tersembunyi. Dokumen digunakan sebagai campuran topik, dimana topik adalah distribusi probabilitas atas kumpulan istilah ini. Setiap dokumen kemudian dilihat sebagai distribusi probabilitas atas serangkaian topik. Dapat dianggap bahwa data berasal dari proses generatif yang didefinisikan oleh distribusi probabilitas gabungan atas apa yang diamati dan apa yang disembunyikan (Hamed Jelodar Y. W., 2017).

3.8.3 Gibbs Sampling

Algoritma Collapsed Gibbs Sampling (CGS) untuk LDA, diperkenalkan oleh Griffiths dan Steyvers (2004), mengesampingkan parameter φ dan θ , dan hanya

beroperasi pada laten tugas variabel z . Ini mengarah pada algoritma *Markov Chain Monte Carlo* (MCMC) yang sederhana namun efektif yang menggabungkan jauh lebih cepat daripada algoritma *gibbs sampling* yang naif. Untuk menghitung pembaruan CGS efisien, algoritma menggunakan beberapa matriks hitungan selama pengambilan sampel. *Gibbs sampling* merupakan cara digunakan untuk mengestimasi $p(z|w)$. Cara kerjanya adalah pertama dengan insialisasi z dengan mengacak topik untuk setiap kata, lalu melakukan sampel z dari semua kata atas distribusi kata dengan topik dari seluruh dokumen dan topik dengan dokumen. Kemudian meng-assign topik hasil sampel untuk setiap kata. :

$$\theta_{dw,k} = \frac{n_{w,k}^w + \beta}{n_{w,k} + W\beta} \frac{n_{w,k}^{dw} + \alpha}{n_w^{dw} + K\alpha} \quad (3.9)$$

dimana :

$\theta_{dw,k}$ = $p(z|w)$ atau probabilitas kata terhadap topik

$n_{w,k}^w$ = banyak kata w di assign ke topik k di setiap dokumen

β = *dirichlet* parameter atas distribusi kata terhadap topik di corpus (semua dokumen)

$n_{w,k}^{dw}$ = banyak topik k di assign ke dokumen d

α = *dirichlet* parameter atas distribusi topik terhadap dokumen

$n_{w,k}$ = banyak kata selain w di assign ke topik k di setiap dokumen

W = jumlah variasi kata di dalam *corpus*

n_w^{dw} = banyak topik selain k di assign ke d

(Setijohatmo et al., 2020)

3.8.4 *Topic Coherence*

Topic coherence mencetak topik dengan mengukur tingkat kesamaan semantik antara kata-kata dengan skor tinggi pada sebuah topik. Pengukuran ini membantu membedakan antara topik yang dapat dimaknai secara semantik dan topik yang hasil interpretasi manusia (Muhammad Habibi, 2020). Untuk evaluasi, dipertimbangkan dua ukuran koherensi baru yang dirancang untuk LDA, yang keduanya telah terbukti cocok dengan penilaian manusia tentang kualitas topik:

- 1) Ukuran UCI
- 2) Ukuran UMass (Mimno et al., 2011)

Kedua ukuran menghitung koherensi suatu topik sebagai jumlah skor kesamaan distribusi berpasangan atas himpunan kata-kata topik, V . Lalu menggeneralisasi ini sebagai :

$$coherence(V) = \sum_{(v_i, v_j) \in V} score(v_i, v_j, \epsilon) \quad (3.8)$$

di mana V adalah himpunan kata yang menjelaskan topik dan ϵ menunjukkan faktor pemulusan yang menjamin bahwa skor mengembalikan bilangan asli, biasanya $V=1$. Kemudian metrik UCI menentukan skor pasangan kata untuk menjadi informasi timbal balik pointwise (PMI) antara dua kata, yaitu

$$score(v_i, v_j, \epsilon) = \log \frac{p(v_i, v_j) + \epsilon}{p(v_i)p(v_j)} \quad (3.9)$$

Probabilitas kata dihitung dengan menghitung frekuensi kemunculan kata bersama di jendela geser melalui korpus eksternal, seperti Wikipedia. Untuk sebagian derajat, metrik ini dapat dianggap sebagai eksternal dibandingkan dengan evaluasi semantik yang diketahui. Kemudian metrik UMass menentukan skor yang akan didasarkan pada kejadian bersama dokumen:

$$score(v_i, v_j, \epsilon) = \log \frac{D(v_i, v_j) + \epsilon}{D(v_j)} \quad (3.10)$$

dimana $D(v_i, v_j)$ menghitung jumlah dokumen yang mengandung kata i dan j dan $D(v_j)$ menghitung jumlah dokumen yang mengandung j . Secara signifikan, metrik UMass menghitung jumlah ini di atas korpus asli yang digunakan untuk melatih model topik, bukan daripada korpus eksternal. Metrik ini lebih bersifat intrinsik. Ini mencoba untuk mengkonfirmasi bahwa model data yang dipelajari diketahui berada di korpus. (Morrison et al., 1993)

BAB IV

METODOLOGI PENELITIAN

4.1 Populasi dan Sampel

Populasi yang digunakan dalam penelitian ini adalah seluruh komentar/tanggapan tentang PPKM di Indonesia melalui media sosial *Twitter*. Sampel penelitian ini yaitu data komentar *tweet* dari tanggal 1 Januari 2021 sampai 30 November 2021 dengan total *tweet* sebanyak 28530 *tweet*.

Penelitian ini menggunakan API sampel akademik 10% yaitu titik akhir yang disediakan secara gratis oleh *Twitter* kepada siapa saja yang ingin mendapatkan sampel 10% dari data *Twitter*. Pengguna API tidak memiliki kemampuan untuk menentukan kueri tertentu ke titik akhir ini dan hanya dapat terhubung serta menerima serangkaian *tweet* tetap dari *Twitter*. Setiap pengguna yang terhubung ke titik akhir ini akan menerima kumpulan *tweet* yang sama. Pada *Twitter* tidak mengungkapkan detail teknis tentang mekanisme pengambilan sampel API Sampel. Cara sampel ini dibuat murni berdasarkan stempel waktu milidetik saat *tweet* tiba di server *Twitter*. Ketika sebuah *tweet* tiba di salah satu server *Twitter*, akan diberi ID dan bagian dari ID ini adalah stempel waktu milidetik saat *tweet* tiba di server (Pfeffer et al., 2018).

4.2 Sumber Data

Sumber data dari penelitian ini adalah data *tweet* mengenai PPKM di Indonesia pada media sosial *Twitter*.

4.3 Variabel dan Definisi Variabel

Variabel yang digunakan dalam penelitian ini dijelaskan pada **Tabel 4.1** berikut:

Tabel 4.1 Variabel Penelitian

| Variabel | Definisi Variabel |
|--------------|---|
| <i>Tweet</i> | Tanggapan atau komentar terkait PPKM di Indonesia yang diambil dari social media <i>Twitter</i> . |

4.4 Metode Pengambilan Data

Metode pengambilan data yang digunakan pada penelitian ini adalah metode *scrapping* pada media sosial *Twitter* menggunakan *Twitter API* dengan kata kunci “PPKM”, “Kebijakan PPKM”, dan “Pemberlakuan Pembatasan Kegiatan Masyarakat” dari tanggal 1 Januari 2021 sampai 30 November 2021 dengan total sebanyak 28530 komentar. Pengambilan data ini menggunakan metode *web scrapping* dengan menggunakan *software Phyton* dan *R studio*.

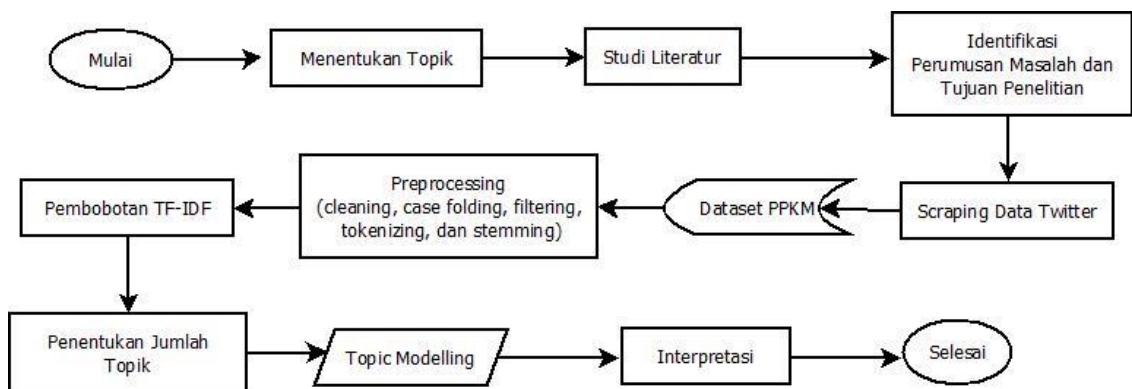
4.5 Metode Analisis Data

Penelitian dilakukan menggunakan bantuan *software* yaitu, *Microsoft excel*, *R studio*, dan *Phyton*. Metode yang digunakan pada penelitian ini adalah :

1. *Text mining* digunakan untuk penambangan data dari kumpulan data teks yang bertujuan untuk menemukan pengetahuan berupa informasi atau pola dari data teks yang tidak terstruktur atau semi terstruktur.
2. Analisis deskriptif untuk menggambarkan secara umum tentang data komentar PPKM di Indonesia.
3. Analisis sentimen untuk melihat sentimen masyarakat terkait kebijakan PPKM apakah positif, negatif, atau netral.
4. TF-IDF digunakan untuk pembobotan kata.
5. Analisis *topic modelling* menggunakan LDA untuk menyimpulkan topik tentang PPKM di Indonesia yang muncul pada *Twitter*.

5.6 Tahapan Penelitian

Tahapan penelitian pada penelitian ini dijelaskan pada gambar *flowchart* dibawah ini.



Gambar 4.1 Diagram Alir Penelitian

Adapun tahapan penelitian berdasarkan **Gambar 4.1** adalah sebagai berikut:

1. Menentukan topik yang akan diangkat dalam penelitian.
2. Melakukan studi literatur terkait topik penelitian guna mencari informasi yang relevan dengan masalah yang akan diteliti.
3. Mengidentifikasi perumusan masalah dan tujuan penelitian yang akan diangkat sebagai bahan penelitian yang dilakukan.
4. Pengambilan data dengan cara *sraping* data di media sosial *Twitter* dengan bantuan API *Twitter*.
5. Didapatkan dataset PPKM dari hasil *sraping* data *Twitter*.
6. Melakukan *preprocessing* data sebagai tahap awal pemrosesan data yang terdiri dari *cleaning*, *case folding*, *filtering*, *tokenizing*, dan *stemming*.
7. Pembobotan TF-IDF untuk memberikan bobot hubungan pada suatu *term* (kata) terhadap suatu dokumen.
8. Menentukan jumlah topik yang digunakan dengan koherensi topik berdasarkan nilai koherensi tertinggi.
9. Didapatkan hasil pemodelan topik menggunakan metode LDA.
10. Interpretasi pemodelan topik yang dihasilkan.

BAB V

HASIL DAN PEMBAHASAN

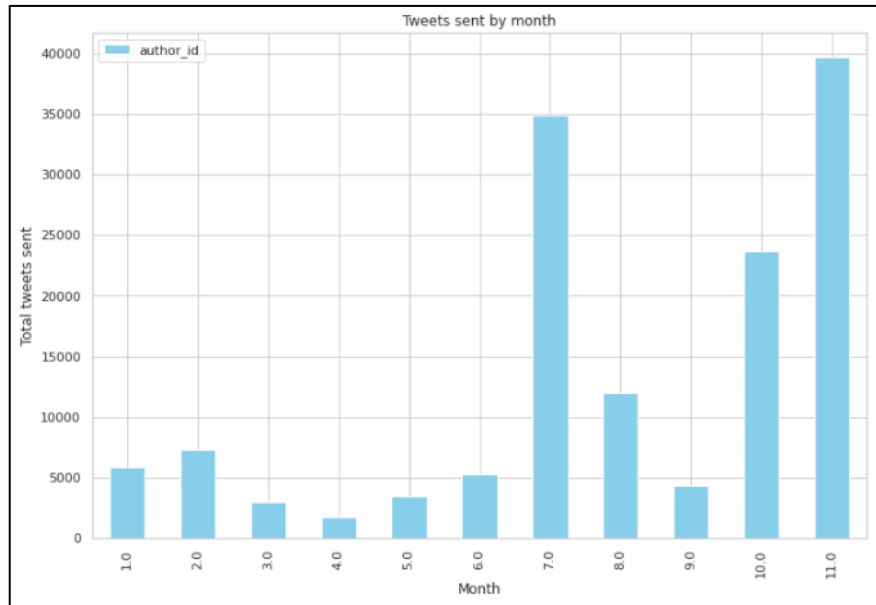
5.1 Analisis Deskriptif

Analisis deskriptif dari penelitian *topic modelling* ini, tahap awal sebelum dilakukan *preprocessing*. Jumlah data yang digunakan dalam penelitian ini berjumlah 28530 baris data *Excel*.

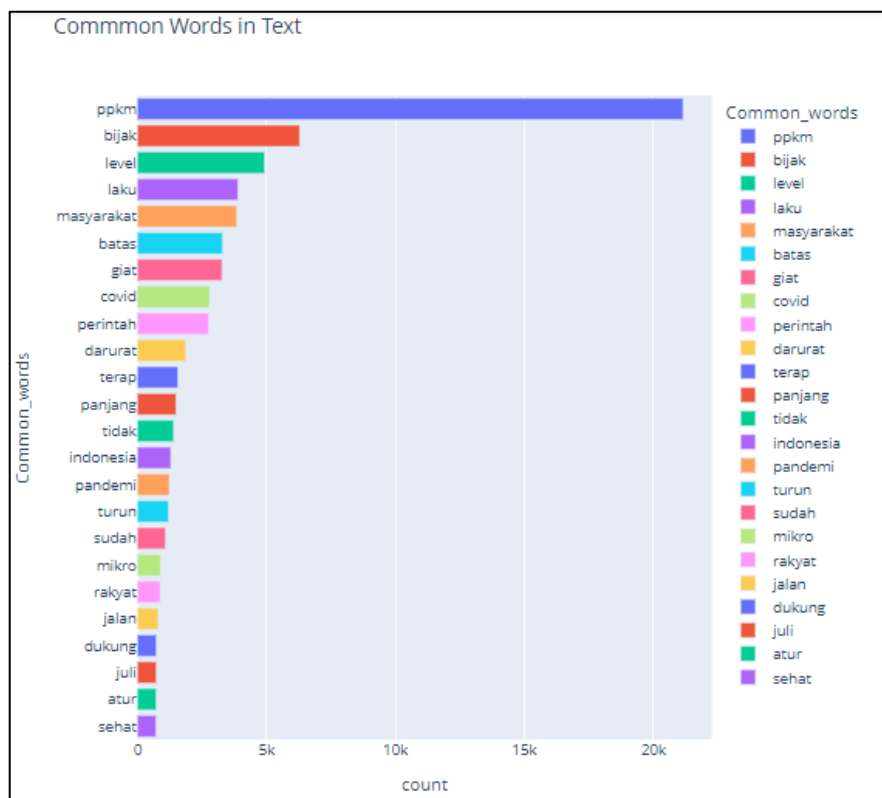
Tabel 5.1 Data PPKM

| No | <i>Tweet</i> |
|----|--|
| 1 | Naik lagi ke level 2, mari tweeps kita tingkatkan kehati-hatian dan perketat proses status ppkm dki jakarta naik level |
| 2 | Pls bgt ini pemerenta kudu ppkm /psbb lagi lah desember s.d januari 2022 minggu ke-1 atau 2 naik tajam varian omicron ga terkendali libur natal taun baru trus w ga kuliah offline |
| 3 | @Aron24996924 Nahh tuh tau. Sebelum PPKM level 3 menjemput |
| 4 | Kabarnya ga boleh cuti libur tahun baru 🙄, Kerja,Kerja, Tipes... harusnya sih PPKM ya srog, Hak libur ya tetap diberikan Pa. hidup pengangguran |
| 5 | Indonesia Tutup Pintu Masuk Bagi Negara Ini !! |

Berdasarkan **Tabel 5.1** merupakan tampilan 5 data *tweet* teratas dari keseluruhan data *tweet* tentang PPKM. *Tweet* tersebut masih terdiri banyak kata-kata yang tidak diperlukan untuk analisis seperti RT, @ RadioElshinta, @Aron24996924, dan lainnya. Sehingga sebelum dilakukan analisis harus dilakukan *cleaning* dan *preprocessing* agar data yang digunakan sesuai. Untuk analisis peneliti membuat 2 model visualisasi, yaitu diagram batang jumlah *tweet* berdasarkan bulan dan diagram batang jumlah kata paling banyak muncul pada data *tweet* PPKM. Berikut pada **Gambar 5.1** dan **Gambar 5.2** merupakan hasil visualisasi data *tweet* tersebut.



Gambar 5.1 Diagram batang jumlah *tweet*



Gambar 5.2 Diagram batang jumlah kata

Pada **Gambar 5.1** menunjukkan diagram batang jumlah *tweet* pada tahun 2021 berdasarkan bulan, yaitu bulan 1.0 (januari), 2.0 (februari), 3.0 (maret), 4.0 (april), 5.0 (mei), 6.0 (juni), 7.0 (juli), 8.0 (agustus), 9.0 (september), 10.0 (oktober), 11.0 (november). Dapat dilihat untuk jumlah *tweet* terbanyak berada pada bulan

November dan Juli. Hal tersebut diperkirakan pada bulan november masyarakat banyak membicarakan tentang PPKM karena adanya pemerintah yang akan menerapkan kebijakan tersebut hingga libur nataru pada bulan Desember-Januari. Sehingga rencana kebijakan tersebut menuai banyak komentar dari masyarakat.

Kemudian dilihat berdasarkan **Gambar 5.2** yang merupakan diagram batang jumlah kata yang muncul pada *tweet* PPKM. Untuk frekuensi kata paling banyak dibicarakan masyarakat diantaranya PPKM, bijak, level, laku, dan lainnya. Dari diagram tersebut dapat dilihat secara garis besar topik pembicaraan masyarakat adalah tentang penerapan PPKM dengan banyak level guna menekan angka kasus covid-19.

5.2 Preprocessing

Data *tweet* yang didapatkan dari *scraping* menggunakan API *Twitter* perlu diproses terlebih dahulu sebelum dapat digunakan untuk analisis pemodelan topik. Sebagian besar data *Twitter* sangat tidak terstruktur. Mungkin ada kesalahan ketik, penggunaan bahasa gaul, dan kesalahan tata bahasa. Kemudian langkah pembersihan diterapkan pada data hasil *scraping* yaitu data PPKM untuk menghasilkan data terstruktur. Terdapat beberapa langkah dalam melakukan *preprocessing* data diantaranya *cleaning*, *case folding*, *filtering*, *tokenizing*, dan *stemming*.

5.2.1 Cleaning

Data yang diperoleh dari *Twitter* biasanya berisi banyak baris baru, URL, tanda baca, angka, sebutan, atau nama pengguna atau entitas HTML seperti `<`; `>`; `&`; yang tertanam dalam data asli. Oleh karena itu, entitas-entitas ini perlu dibersihkan. Berikut pada **Tabel 5.2** merupakan tampilan sebagian data hasil *cleaninig*.

Tabel 5.2 Hasil *Cleaning*

| Komentar Sebelum <i>Cleaning</i> | Komentar Setelah <i>Cleaning</i> |
|--|---|
| Naik lagi ke level 2, mari tweeps kita tingkatkan kehati-hatian dan perketat prokes status ppkm dki jakarta naik level | Naik lagi ke level mari tweeps kita tingkatkan kehati-hatian dan perketat prokes status ppkm dki jakarta naik level |

Hasil data *tweet* sebelum dan setelah *cleaning* dapat dilihat pada **Tabel 5.2** dimana data sebelum dilakukan *cleaning* masih terdapat entitas seperti nama pengguna, angka-angka yang dianggap tidak perlu, dan entitas lainnya yang tidak diperlukan akan dihilangkan sehingga nampak pada kolom *tweet* sesudah untuk hasil yang telah dilakukan *cleaning*.

5.2.2 Case Folding

Pada data hasil *scraping* tidak semua memiliki penggunaan huruf kapital yang konsisten dan teratur. *Case folding* berguna untuk menyamaratakan arti kata sehingga arti dalam kata tidak berubah dengan mengganti semua huruf dalam kata menjadi huruf kecil semua. Karakter selain huruf dihilangkan dan dianggap sebagai batas pemisah untuk urutan satu atau lebih karakter yang digunakan. Hasil dari *case folding* sebagai berikut.

Tabel 5.3 Hasil *Case Folding*

| Komentar Sebelum <i>Case folding</i> | Komentar Setelah <i>Case folding</i> |
|---|---|
| Naik lagi ke level mari tweeps kita tingkatkan kehati-hatian dan perketat prokes status ppkm dki jakarta naik level | naik lagi ke level mari tweeps kita tingkatkan kehati-hatian dan perketat prokes status ppkm dki jakarta naik level |

Pada **Tabel 5.3** menunjukkan komentar sebelum dan setelah *case folding*. Terlihat komentar sebelum dilakukan *case folding* masih terdapat huruf kapital dan setelah dilakukan *case folding* komentar sudah berupa huruf kecil semua.

5.2.3 Filtering

Tahap *filtering* adalah menyaring beberapa kata dari dokumen, yang disebut sebagai *stopwords*. Kata-kata ini biasanya kata ganti, artikel, atau beberapa yang sangat umum kata-kata yang dikeluarkan dari kosa kata karena memiliki sedikit nilai dalam pencarian informasi teks. Contoh *stopwords* disini adalah “di”, “ke”, “dan”, “dari” dan lainnya.

Tabel 5.4 Hasil *Filtering*

| Komentar Sebelum <i>Filtering</i> | Komentar Setelah <i>Filtering</i> |
|---|--|
| naik lagi ke level mari tweeps kita tingkatkan kehati-hatian dan perketat prokes status ppkm dki jakarta naik level | naik level mari kita tingkatkan kehati-hatian perketat prokes status ppkm dki jakarta naik level |

Terlihat pada **Tabel 5.4** dilakukan penghapusan kata-kata seperti “kembali” dan “hingga” karena kata-kata ini dianggap tidak dibutuhkan dalam analisis. Sehingga untuk analisis selanjutnya kata-kata tersebut tidak digunakan.

5.2.4 *Tokenizing*

Tokenizing yaitu proses untuk pemisahan frasa, kalimat, paragraf, atau seluruh dokumen teks menjadi masing-masing kata atau individual. Masing-masing unit yang lebih kecil ini disebut token. Hasil *tokenizing* dapat dilihat pada **Tabel 5.5** berikut.

Tabel 5.5 Hasil *Tokenizing*

| Komentar Sebelum <i>Tokenizing</i> | Komentar Setelah <i>Tokenizing</i> |
|--|--|
| naik level mari kita tingkatkan kehati-hatian perketat prokes status ppkm dki jakarta naik level | 'naik', 'level', 'mari', 'kita', 'tingkatkan', 'kehati', 'hatian', 'perketat', 'prokes', 'status', 'ppkm', 'dki', 'jakarta', 'naik', 'level' |

Dapat terlihat pada **Tabel 5.5** kata-kata yang memiliki maknanya sudah dikutip kutip setiap *term* proses akan dilanjutkan perhitungan *Term Frequency Invers Document Frequency* (TF-IDF).

5.2.5 *Stemming*

Beberapa kata dalam komentar memiliki arti yang sama tetapi dalam bentuk yang berbeda. *Stemming* disini merupakan proses menghilangkan imbuhan dari kata untuk mengubah kata menjadi bentuk dasarnya.

Tabel 5.6 Hasil *Stemming*

| Komentar Sebelum <i>Stemming</i> | Komentar Setelah <i>Stemming</i> |
|--|--|
| 'naik', 'level', 'mari', 'kita', 'tingkatkan', 'kehati', 'hatian', 'perketat', 'prokes', 'status', 'ppkm', 'dki', 'jakarta', 'naik', 'level' | 'naik', 'level', 'tingkat', 'hati', 'hati', 'ketat', 'prokes', 'status', 'ppkm', 'dki', 'jakarta', 'naik', 'level' |

Sebagaimana pada **Tabel 5.6** terlihat untuk perubahan kata-kata dalam komentar menjadi kata dasar seperti kata sebelum dilakukan *stemming* contohnya “tingkatkan” dan setelah dilakukan *stemming* berubah menjadi “tingkat”. Begitupun untuk kata-kata yang lainnya. Hasil *stemming* diperoleh data yang digunakan untuk analisis selanjutnya yaitu 28375 baris data.

5.3 Analisis Sentimen

Pada analisis sentimen disini dilakukan pelabelan text data *tweet* PPKM yang memuat informasi subjektif dalam sebuah ekspresi, yaitu opini, penilaian, emosi, atau sikap terhadap topik, orang, atau entitas dimana akan diklasifikasikan sebagai positif, negatif, atau netral. Menggunakan data hasil *preprocessing* pada data *tweet* lalu ditentukan polaritas sentimen *tweet* dengan *InSet Lexicon* atau *Indonesian Sentiment Lexicon* (<https://github.com/rifyahmadsaputra/Sentiment-Analysis-Online-Lectures-in-Indonesia/tree/main/data>) . *Lexicon* yang digunakan terdiri dari 3597 kata positif dan 6599 kata negatif berbahasa Indonesia dimana masing-masing kata telah memiliki bobot nilai atau *polarity score* untuk mengklasifikasikan jenis sentimen. *Polarity score* pada setiap katanya berkisar dari -5 sampai +5. Contoh dari kata negatif dan positif beserta bobotnya seperti berikut.

Tabel 5.7 Contoh daftar kata *lexicon* positif

| | |
|-----------|---|
| hai | 3 |
| merekam | 2 |
| ekstensif | 3 |
| paripurna | 1 |
| detail | 2 |

Tabel 5.8 Contoh daftar kata *lexicon* negatif

| | |
|--------|----|
| ngga | -2 |
| borong | -1 |
| lever | -2 |
| kasian | -3 |
| gamau | -4 |

Berdasarkan bobot nilai sesuai contoh daftar pada **Gambar 5.7** dan **Gambar 5.8**, menggunakan data *tweet* PPKM ini dilakukan contoh penghitungan *polarity score* dan *polarity* pada sentimen netral, negatif, dan positif pada masing-masing *tweet* berikut.

Tabel 5.9 Contoh Perhitungan Sentimen Netral

| Kata | <i>Polarity_Score</i> | <i>Polarity</i> |
|------------|-----------------------|-----------------|
| ['tau'] | 0 | neutral |
| ['ppkm'] | 0 | neutral |
| ['level'] | 0 | neutral |
| ['jemput'] | 0 | neutral |
| Pelabelan | 0 | neutral |

Tabel 5.10 Contoh Perhitungan Sentimen Negatif

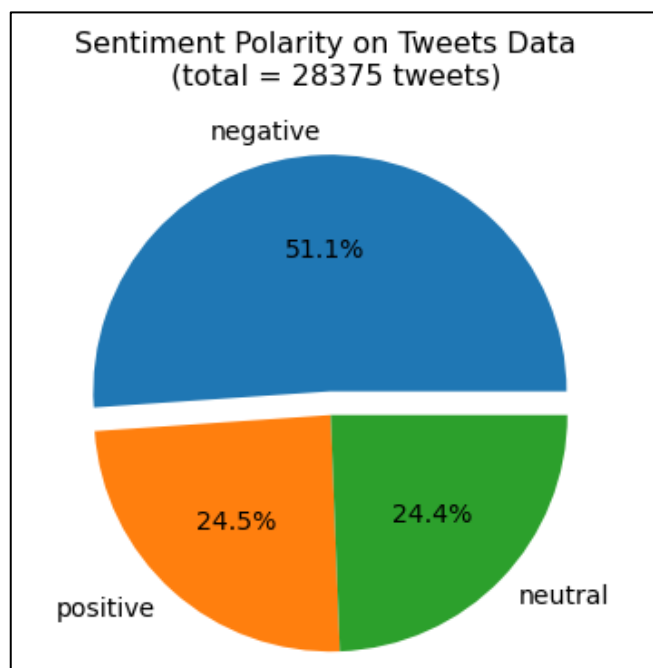
| Kata | <i>Polarity_Score</i> | <i>Polarity</i> |
|--|-----------------------|-----------------|
| ['kabar'] | -1 | negative |
| ['tidak'] | -5 | negative |
| ['cuti'] | 1 | positive |
| ['libur'] | 3 | positive |
| Pelabelan (Jumlah Positif + Jumlah Negatif) | -2 | negative |

Tabel 5.11 Contoh Perhitungan Sentimen Positif

| Kata | <i>Polarity_Score</i> | <i>Polarity</i> |
|--------------|-----------------------|-----------------|
| ['perintah'] | 0 | neutral |
| ['panjang'] | 2 | positive |
| ['ppkm'] | 0 | neutral |
| ['jawabali'] | 0 | neutral |
| ['desember'] | 0 | neutral |

| Kata | <i>Polarity_Score</i> | <i>Polarity</i> |
|--|-----------------------|-----------------|
| Pelabelan (Jumlah Positif + Jumlah Negatif) | 2 | positive |

Setelah dilakukan penghitungan *polarity* pada masing-masing *tweet*, dilakukan visualisasi menggunakan *pie chart* untuk melihat besar persentase masing-masing *sentiment polarity* seperti pada **Gambar 5.3** berikut.



Gambar 5.3 Persentase sentimen data *tweet* PPKM

Pie chart menunjukkan persentase *sentiment polarity* pada data *tweet* PPKM dengan total *tweet* sebanyak 28375 *tweets*. Hasil tersebut dapat dilihat bahwa sentimen negatif memiliki persentase terbesar dibandingkan sentimen positif dan netral dengan nilai persentase sebesar 51,1% atau sebanyak 14496 *tweets*. Sedangkan untuk sentimen positif sebesar 24,5 % atau 6959 *tweets* dan sentimen netral sebesar 24,4% atau 6920 *tweets*. Analisis sentimen ini menyimpulkan bahwa untuk komentar tentang PPKM di *Twitter* kebanyakan masyarakat tidak mendukung ataupun menolak tentang adanya kebijakan ini.

5.4 *Term Frequency Invers Document Frequency (TF-IDF)*

Data hasil *preprocessing* kemudian dilakukan pembobotna TF-IDF yaitu untuk mengevaluasi seberapa relevan sebuah kata dengan dokumen dalam kumpulan dokumen yang dianalisis. TF-IDF untuk sebuah kata dalam dokumen

dihitung dengan mengalikan dua metrik yang berbeda yaitu TF (*Term Frequency*) atau frekuensi kata dalam dokumen dengan IDF (*Invers Document Frequency*) atau kebalikan frekuensi dokumen dari kata dalam sebuah dokumen.

Tabel 5.12 Contoh Data Pembobotan *TF-IDF*

| No | Token |
|----|--|
| 0 | perintah panjang atur laku batas giat masyarakat ppkm jawa bali desember |
| 1 | level tingkat hati-hati ketat prokes status ppkm dki jakarta |
| 2 | renta ppkm psbb desember sihada januari minggu tajam |
| 3 | jakarta bebas pandemi masuk kategori ppkm level kondisi natal turun level dki |
| 4 | status ppkm dki jakarta level |

Berdasarkan data *tweet* PPKM terbentuk hasil nilai TF-IDF dengan matriks 28375 baris dan 19525 kolom, frekuensi kata tertinggi yang banyak digunakan publik dalam *tweet* yaitu kata pemerintah, ppkm, memperpanjang, dan lainnya. Pada **Tabel 5.12** merupakan contoh data untuk penghitungan bobot TF-IDF dimana terdapat 5 komentar data tentang PPKM. Langkah pertama adalah melakukan penghitungan untuk TF sebagaimana **Tabel 5.13** berikut :

Tabel 5.13 Contoh Perhitungan *TF*

| Token | TF | | | | |
|----------|----|----|----|----|----|
| | D0 | D1 | D2 | D3 | D4 |
| atur | 1 | | | | |
| bali | 1 | | | | |
| batas | 1 | | | | |
| bebas | 1 | | | | |
| desember | 1 | | 1 | | |
| dki | | 1 | | 1 | 1 |
| januari | | | 1 | | |
| kondisi | | | | 1 | |
| level | | 1 | | 2 | 1 |
| ppkm | 1 | 1 | 1 | 1 | 1 |

Hasil contoh penghitungan TF sesuai pada **Tabel 5.13** menunjukkan banyaknya kata yang muncul pada masing-masing dokumen seperti kata “atur” muncul 1 kali pada dokumen 0, kata “bali” muncul 1 kali pada dokumen 0, dan lainnya. Lalu melakukan penghitungan DF seperti berikut :

Tabel 5.14 Contoh Perhitungan *DF*

| Token | TF | | | | | DF |
|----------|----|----|----|----|----|----|
| | D0 | D1 | D2 | D3 | D4 | |
| atur | 1 | | | | | 1 |
| bali | 1 | | | | | 1 |
| batas | 1 | | | | | 1 |
| bebas | 1 | | | | | 1 |
| desember | 1 | | 1 | | | 2 |
| dki | | 1 | | 1 | 1 | 3 |
| januari | | | 1 | | | 1 |
| kondisi | | | | 1 | | 1 |
| level | | 1 | | 2 | 1 | 4 |
| ppkm | 1 | 1 | 1 | 1 | 1 | 5 |

Berdasarkan informasi **Tabel 5.14** dimana nilai DF merupakan banyaknya dokumen yang muncul suatu *term* tertentu seperti kata “atur” hanya muncul pada dokumen 0 dari keseluruhan dokumen, kata “ppkm” muncul pada kelima dokumen sehingga jumlah DF disini sebanyak 5. Kemudian menghitung bobot TF-IDF menggunakan hasil TF dan DF yang telah didapatkan sebelumnya sebagaimana pada **Tabel 5.15** berikut :

Tabel 5.15 Contoh Perhitungan TF-IDF

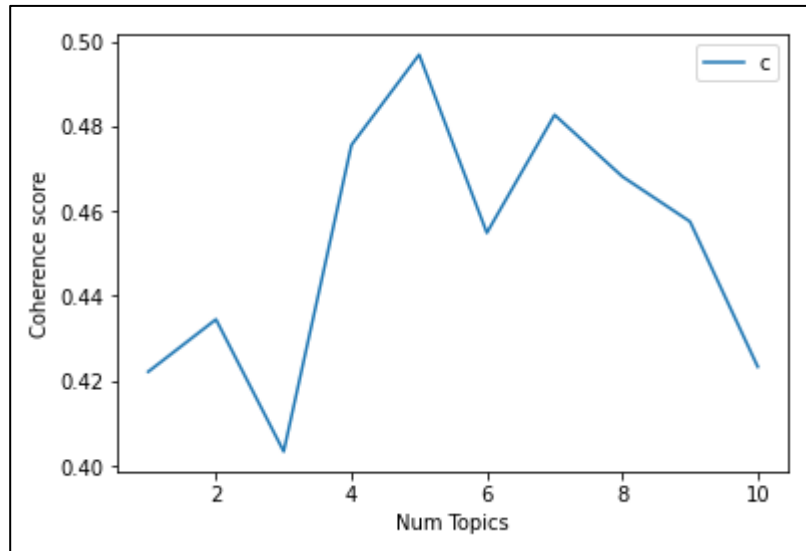
| Token | TF | | | | | DF | D/DF | IDF | TF-IDF |
|-------|----|----|----|----|----|----|-------|------------------|--------|
| | D0 | D1 | D2 | D3 | D4 | | | | |
| atur | 1 | | | | | 1 | 5/1=5 | $\ln(5)+1=2,609$ | 2,609 |
| bali | 1 | | | | | 1 | 5/1=5 | $\ln(5)+1=2,609$ | 2,609 |
| batas | 1 | | | | | 1 | 5/1=5 | $\ln(5)+1=2,609$ | 2,609 |
| bebas | 1 | | | | | 1 | 5/1=5 | $\ln(5)+1=2,609$ | 2,609 |

| Token | TF | | | | | DF | D/DF | IDF | TF-IDF |
|----------|----|----|----|----|----|----|------------|---------------------|--------|
| | D0 | D1 | D2 | D3 | D4 | | | | |
| desember | 1 | | 1 | | | 2 | $5/2=2,5$ | $\ln(2,5)+1=1,916$ | 3,833 |
| dki | | 1 | | 1 | 1 | 3 | $5/3=1,67$ | $\ln(1,67)+1=1,511$ | 4,532 |
| januari | | | 1 | | | 1 | $5/1=5$ | $\ln(5)+1=2,609$ | 2,609 |
| kondisi | | | | 1 | | 1 | $5/1=5$ | $\ln(5)+1=2,609$ | 2,609 |
| level | | 1 | | 2 | 1 | 3 | $5/3=1,67$ | $\ln(1,67)+1=1,511$ | 4,532 |
| ppkm | 1 | 1 | 1 | 1 | 1 | 5 | $5/5=1$ | $\ln(1)+1=1$ | 5 |

Pada **Tabel 5.15** dilakukan penghitungan D/DF terlebih dahulu untuk dapat mencari nilai IDF. Contohnya untuk hasil IDF dari kata “atur” dihitung dengan $D=5$ yaitu banyaknya dokumen sebanyak 5 dibagi dengan DF sebesar 1 sehingga didapatkan nilai IDF yaitu 2,609. Lalu untuk nilai TF-IDF dihitung dari hasil perkalian IDF dengan D/DF sehingga menghasilkan TF-IDF sebesar 2,609.

5.5 Hasil *Topic Modelling* dengan Bigram Trigram

Topic modelling dengan metode LDA mengklasifikasikan atau mengkategorikan teks ke dalam dokumen dan kata-kata per topik. Pada penelitian ini didapatkan pemodelan topik yang mengkategorikan komentar publik tentang PPKM yang merupakan hasil *tweet* di sosial media *Twitter* ke dalam beberapa topik untuk mengetahui topik yang paling sering muncul yang berkaitan dengan PPKM. Dalam menentukan banyaknya topik model LDA, pertama menggunakan cara rasional *trial-error* dengan limit maksimum topik 11, 21, dan 31 didapatkan jumlah topik kurang dari 10 topik. Setelah dilakukan dengan percobaan beberapa kali hasil topik yg terbentuk banyak muncul pada jumlah topik 5. Kemudian dengan grafik *coherence score* akan melihat *score* tertinggi sebagai nilai optimum untuk jumlah topik yang akan digunakan seperti pada **Gambar 5.5**.



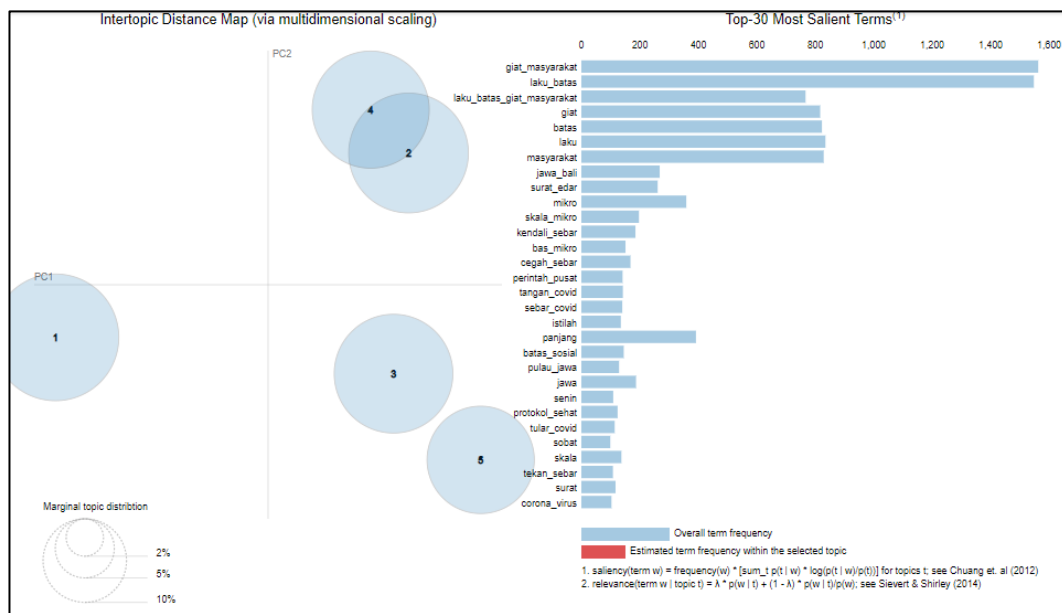
Gambar 5.4 Grafik *Coherence Score* Data PPKM dengan Bigram Trigram

Terlihat dari grafik pada **Gambar 5.4** bahwa nilai koherensi berfluktuasi dari 1 topik sampai 3 topik. Lalu meningkat pesat mencapai 5 topik, dan nilai koherensi mengalami penurunan untuk 6 topik. Namun nilai koherensi yang diperoleh terus menurun hingga 10 topik dan mencapai skor tertinggi pada 5 topik. Hasil ini kemudian dijadikan acuan untuk analisis lebih lanjut, sehingga pembahasan mendalam akan terfokus pada 5 topik. Besarnya nilai koherensi berdasarkan grafik *coherence score* **Gambar 5.4** ditunjukkan pada **Tabel 5.16**

Tabel 5.16 Nilai Koherensi Topik PPKM 5 dengan Bigram Trigram

| <i>Num Value</i> | <i>Coherence Value</i> |
|------------------|------------------------|
| 1 | 0.422071 |
| 2 | 0.434404 |
| 3 | 0.403269 |
| 4 | 0.475496 |
| 5 | 0.496818 |
| 6 | 0.454812 |
| 7 | 0.482622 |
| 8 | 0.468097 |
| 9 | 0.457468 |
| 10 | 0.423241 |

Jumlah topik 5 memiliki peluang terbesar yaitu 0,496818 sehingga ini yang akan dijadikan acuan pemodelan topik. Kemudian dilakukan visualiasi untuk 5 topik yang terbentuk tersebut. Pemodelan topik akan sedikit sulit dipahami jika hanya dengan melihat kombinasi kata dan angka dari model yang terbentuk. Salah satu cara paling efektif untuk memahami data adalah melalui visualisasi *PyLDAvis*. Dengan begitu dapat memungkinkan untuk menginterpretasikan topik dalam model topik seperti diagram pada **Gambar 5.5**



Gambar 5.5 Visualisasi umum *Topic Modelling* PPKM dengan Bigram Trigram

Pada **Gambar 5.5** menunjukkan diagram terbagi menjadi 2 panel yaitu panel kiri dan panel kanan. Panel kiri dengan dimensi gelembung menggambarkan jarak antar topik dan area gelembung sesuai dengan prevalensi relatif topik di dalam korpus. Setiap gelembung mewakili sebuah topik dimana semakin besar gelembung, semakin tinggi persentase jumlah *tweet* dalam korpus tentang topik tersebut. Namun dilihat dari gelembung yang muncul pada 5 topik yang terbentuk menunjukkan bahwa area gelembung memiliki besaran yang cukup sama. Sehingga istilah frekuensi antara topik satu dengan lainnya tidak berbeda jauh bobotnya. Sedangkan panel kanan dijelaskan dengan menggunakan diagram batang sebagai banyaknya frekuensi kata atau istilah yang sering muncul. Dapat dilihat terdapat 30 istilah yang paling relevan untuk topik tertentu. Dari panel kanan hasil pemodelan topik diketahui diagram batang bahwa “giat_masyarakat”, “laku_batas”,

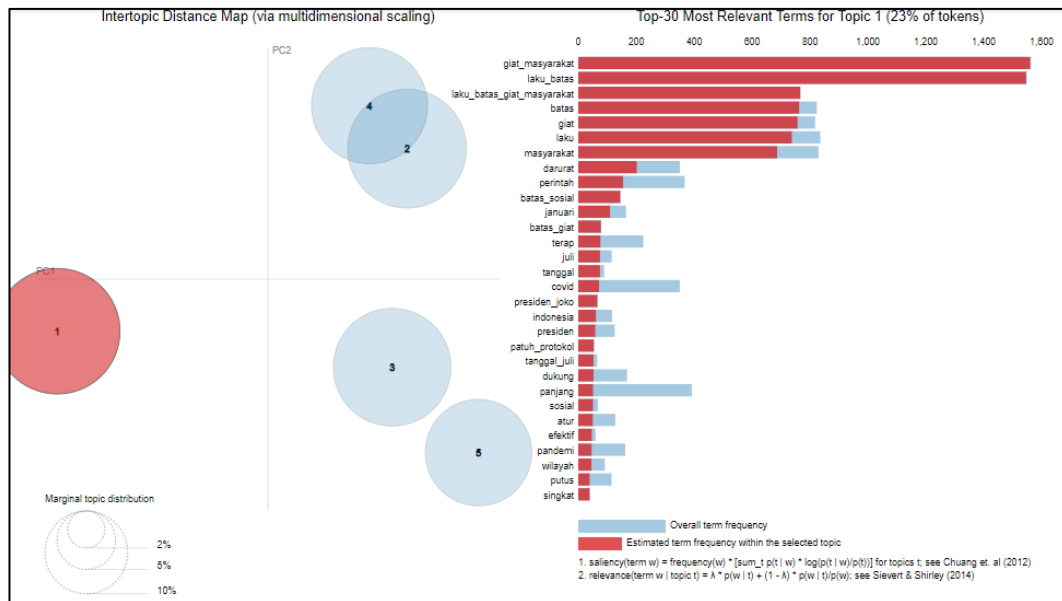
“laku_batas_giat_masyarakat” adalah istilah yang sering dibahas masyarakat ketika berbicara tentang PPKM. Dari 5 topik gelembung, terdapat topik yang beririsan yaitu topik 2 beririsan dengan topik 4. Kemudian memvisualisasikan istilah topik dengan *word cloud* memberikan gambaran kemungkinan adanya istilah dalam topik seperti pada **Gambar 5.6**



Gambar 5.6 *Word cloud* Topik 1 dengan Bigram Trigram

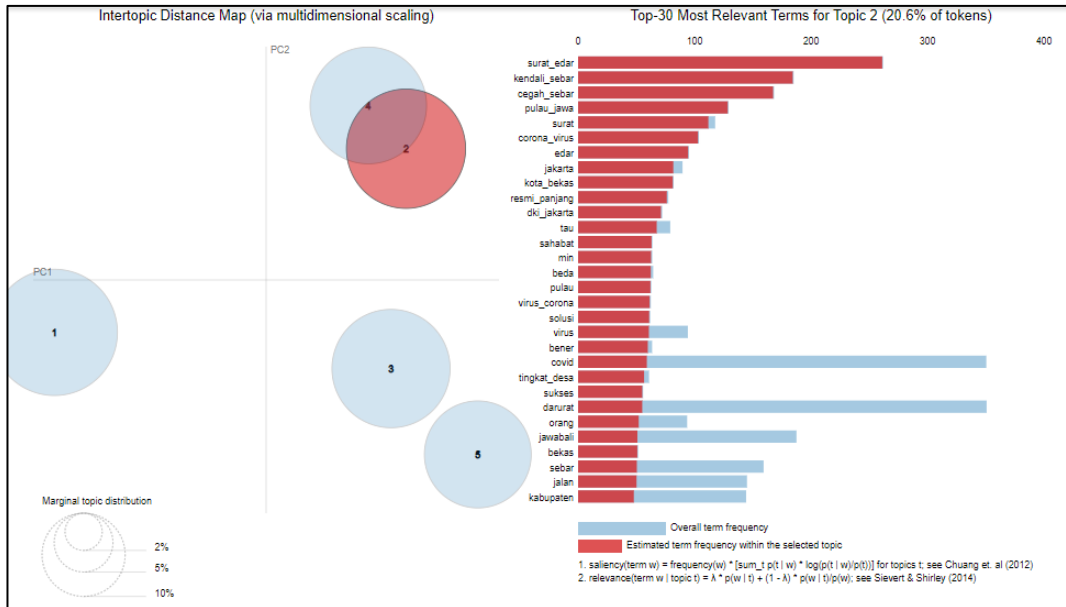
Word cloud disini merupakan kumpulan kata-kata yang digambarkan dalam berbagai ukuran. Visualisasi *word cloud* memberikan gambaran tentang kata-kata yang sering digunakan dalam komentar publik tentang PPKM. Semakin besar ukuran dari kata tersebut maka semakin sering kata tersebut dibicarakan. Dari *wordcloud* tersebut dapat dilihat kata-kata apa paling signifikan yang membangun segmen tertentu dari model yang sedang dibangun. **Gambar 5.6** menunjukkan *wordcloud* topik pertama tentang PPKM dimana “perintah_pusat”, “mikro”, “bas_mikro”, “level” merupakan kata yang banyak dibicarakan publik saat berbicara tentang PPKM. Selanjutnya untuk *wordcloud* yang lain dapat dilihat pada bagian lampiran.

Sebelumnya telah dilakukan visualisasi dengan *PyLDAvis* pada data *tweet* PPKM secara keseluruhan. Lalu untuk melihat frekuensi kata yang muncul dari masing-masing topik dapat dilakukan visualisasi pada topik 1 sampai topik 5 dimana setiap gelembung mewakili sebuah topik. Model topik yang baik akan memiliki gelembung besar dan tidak tumpang tindih yang tersebar di seluruh bagan. Seperti yang dapat dilihat dari grafik pada **Gambar 5.7** untuk hasil visualisasi topik 1.



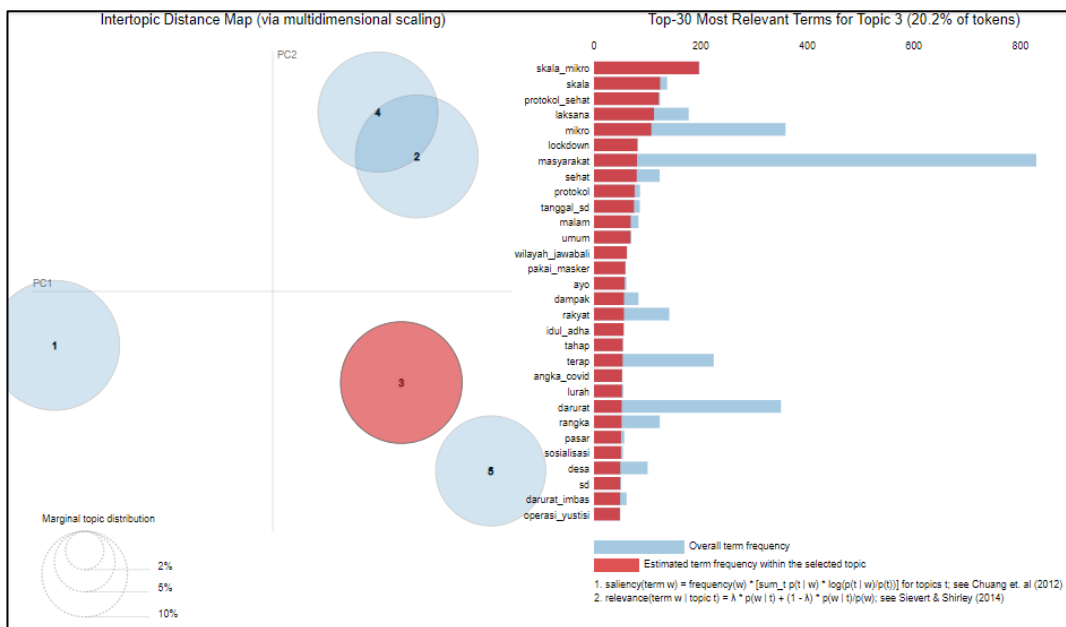
Gambar 5.7 Visualisasi Topik 1 dengan Bigram Trigram

Pada **Gambar 5.7** diagram batang berwarna merah memberikan perkiraan banyaknya istilah tertentu dihasilkan oleh topik tertentu. Seperti hasil diagram topik 1 ada sekitar hampir 1600 kata “giat_masyarakat”, dan istilah ini digunakan hampir 1600 kali dalam topik 1. Kata dengan bilah merah terpanjang adalah kata yang paling banyak digunakan oleh *tweet* topik tersebut. Pada topik 1, menunjukkan bahwa kata diurutkan berdasarkan frekuensi terbanyak adalah giat_masyarakat, laku_batas, laku_batas_giat_masyarakat, batas, giat, laku, masyarakat, darurat, pemerintah, batas_sosial, januari, batas_giat, terap, juli, tanggal, covid, presiden_joko, indonesia, presiden, patuh_protokol, tanggal_juli, dukung, panjang, sosial, atur, efektif, pandemi, wilayah, putus, singkat.



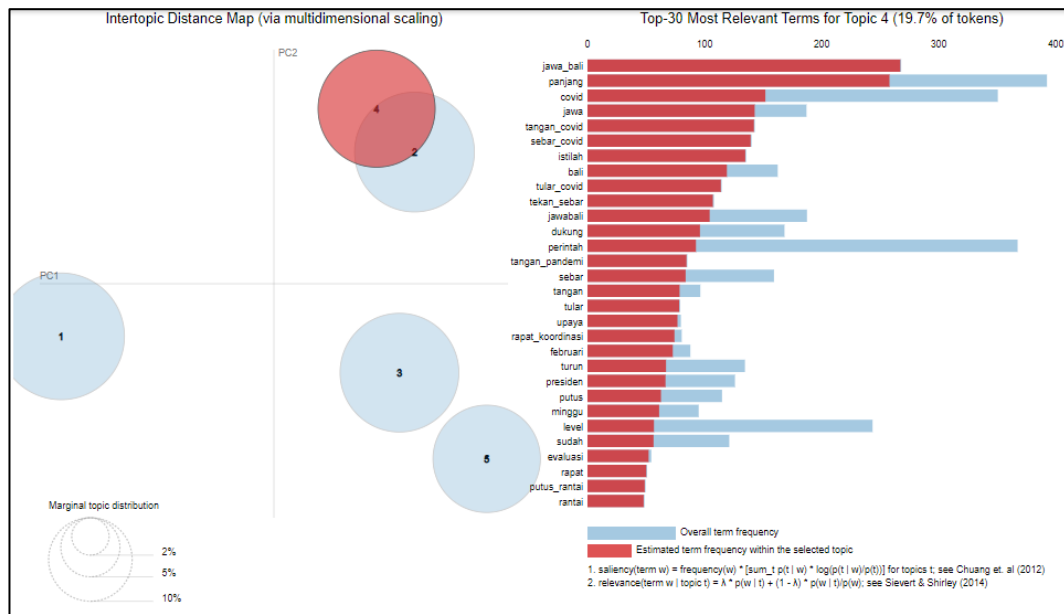
Gambar 5.8 Visualisasi Topik 2 dengan Bigram Trigram

Pada **Gambar 5.8** yang merupakan visualisasi topik 2, menunjukkan bahwa kata diurutkan berdasarkan kata frekuensi terbanyak adalah surat_edar, kendali_sebar, cegah_sebar, pulau_jawa, surat, corona_virus, edar, jakarta, kota_bekas, resmi_panjang, dki_jakarta, tau, sahabat, min, beda, pulau, virus_corona, solusi, virus, bener, covid, tingkat_desa, sukses, darurat, orang, jawabali, bekas, sebar, jalan, kabupaten.



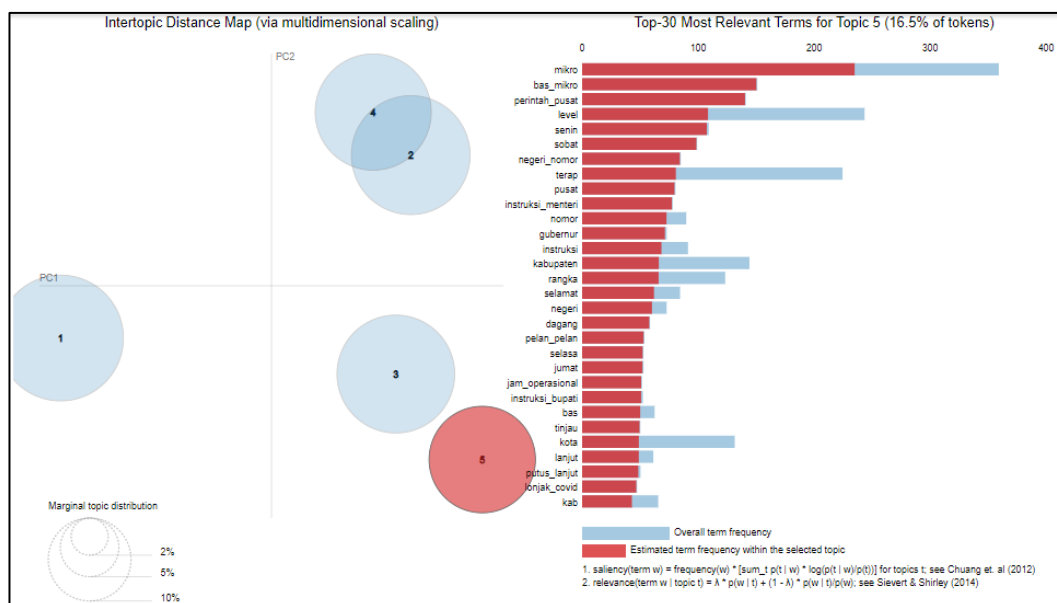
Gambar 5.9 Visualisasi Topik 3 dengan Bigram Trigram

Pada **Gambar 5.9** yang merupakan visualisasi topik 3, menunjukkan bahwa kata diurutkan berdasarkan kata frekuensi terbanyak adalah skala_mikro, skala, protokol_sehat, laksana, mikro, lockdown, masyarakat, sehat, protokol, tanggal_sd, malam, umum, wilayah_jawa, bali, pakai_masker, ayo, dampak, rakyat, idul_adha, tahap, terap, angka_covid, lurah, darurat, rangka, pasar, sosialisasi, desa, sd, darurat_imbas, operasi_yustisi.



Gambar 5.10 Visualisasi Topik 4 dengan Bigram Trigram

Pada **Gambar 5.10** yang merupakan visualisasi topik 4, menunjukkan bahwa kata diurutkan berdasarkan kata frekuensi terbanyak adalah jawa_bali, panjang, covid, jawa, tangan_covid, sebar_covid, istilah, bali, tular_covid, tekan_sebar, jawabali, dukung, perintah, tangan_pandemi, sebar, tangan, tular, upaya, rapat_koordinasi, februari, turun, presiden, putus, minggu, level, sudah, evaluasi, rapat, putus_rantai, rantai



Gambar 5.11 Visualisasi Topik 5 dengan Bigram Trigram

Pada **Gambar 5.11** yang merupakan visualisasi topik 5, menunjukkan bahwa kata diurutkan berdasarkan kata frekuensi terbanyak adalah mikro, bas_mikro, perintah_pusat, level, senin, sobat, negeri_nomor, terap, pusat, instruksi_menteri, nomor, gubernur, instruksi, kabupaten, rangka, selamat, negeri, dagang, pelan_pelan, selasa, jumat, jam_operasional, instruksi_bupati, bas, tinjau, kota, lanjut, putus_lanjut, lonjak_covid, kab.

Kemudian setelah dilakukan visualisasi, dapat dilihat untuk hasil pemodelan topik dari 5 topik yang digunakan tentang perspektif publik tentang PPKM seperti berikut:

Tabel 5.17 Hasil Model Topik dengan Bigram Trigram

| Topik | Model | Kesimpulan |
|-------|---|----------------------------|
| 0 | $0.024 * \text{"mikro"} + 0.016 * \text{"bas_mikro"} + 0.014 * \text{"perintah_pusat"} + 0.011 * \text{"level"} + 0.011 * \text{"senin"} + 0.010 * \text{"sobat"} + 0.009 * \text{"negeri_nomor"} + 0.008 * \text{"terap"} + 0.008 * \text{"pusat"} + 0.008 * \text{"instruksi_menteri"}$ | Instruksi PPKM mikro |
| 1 | $0.022 * \text{"surat_edar"} + 0.015 * \text{"kendali_sebar"} + 0.014 * \text{"cegah_sebar"} + 0.011 * \text{"pulau_jawa"} + 0.009 * \text{"surat"} + 0.009 * \text{"corona_virus"}$ | Surat edar perpanjang PPKM |

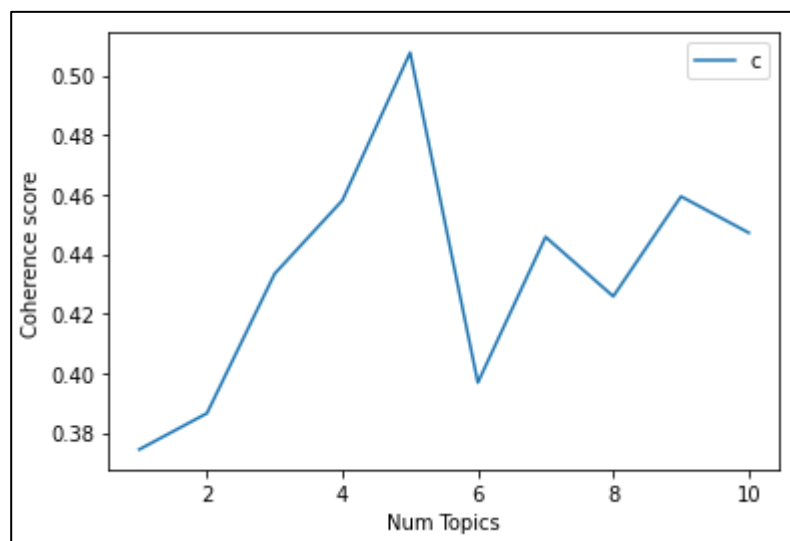
| Topik | Model | Kesimpulan |
|-------|---|---------------------------------------|
| | 0.008*"edar" + 0.007*"jakarta" + 0.007*"kota_bekas" + 0.006*"resmi_panjang" | |
| 2 | 0.017*"skala_mikro" + 0.010*"skala" + 0.010*"protokol_sehat" + 0.009*"laksana" + 0.009*"mikro" + 0.007*"lockdown" + 0.007*"masyarakat" + 0.007*"sehat" + 0.006*"protokol" + 0.006*"tanggal_sd" | Protokol kesehatan saat PPKM mikro |
| 3 | 0.023*"jawa_bali" + 0.022*"panjang" + 0.013*"covid" + 0.012*"jawa" + 0.012*"tangan_covid" + 0.012*"sebar_covid" + 0.012*"istilah" + 0.010*"bali" + 0.010*"tular_covid" + 0.009*"tekan_sebar" | PPKM Jawa Bali |
| 4 | 0.116*"giat_masyarakat" + 0.115*"laku_batas" + 0.057*"laku_batas_giat_masyarakat" + 0.057*"batas" + 0.056*"giat" + 0.055*"laku" + 0.051*"masyarakat" + 0.015*"darurat" + 0.011*"perintah" + 0.011*"batas_sosial" | PPKM darurat |

Pada hasil model **Tabel 5.17** yang terbentuk dengan melihat visualisasi menggunakan *pyLDavis* terdapat beberapa model yang beririsan dimana pada model yang beririsan tersebut ada beberapa kata yang sama dibicarakan masyarakat terkait kebijakan PPKM yang diberlakukan pemerintah. Seperti disini untuk hasil model 2 dan model 4 beririsan yaitu keduanya membicarakan tentang PPKM untuk menekan penyebaran covid-19. Kata yang sama banyak dibicarakan disini seperti kata tekan_sebar, covid, sebar_covid, dan lainnya.

5.6 Hasil Topic Modelling dengan Unigram

Topic modelling dengan metode LDA mengklasifikasikan atau mengkategorikan teks ke dalam dokumen dan kata-kata per topik. Pada penelitian ini didapatkan pemodelan topik yang mengkategorikan komentar publik tentang

PPKM yang merupakan hasil *tweet* di sosial media *Twitter* ke dalam beberapa topik untuk mengetahui topik yang paling sering muncul yang berkaitan dengan PPKM. Sama seperti sebelumnya, dalam menentukan banyaknya topik model LDA, pertama menggunakan cara rasional *trial-error* dengan limit maksimum topik 11, 21, dan 31 didapatkan jumlah topik kurang dari 10 topik. Setelah dilakukan dengan percobaan beberapa kali hasil topik yg terbentuk banyak muncul pada jumlah topik 5. Kemudian dengan grafik *coherence score* akan melihat *score* tertinggi sebagai nilai optimum untuk jumlah topik yang akan digunakan seperti pada pada **Gambar 5.12** berikut :



Gambar 5.12 Grafik *Coherence Score* Data PPKM dengan Unigram

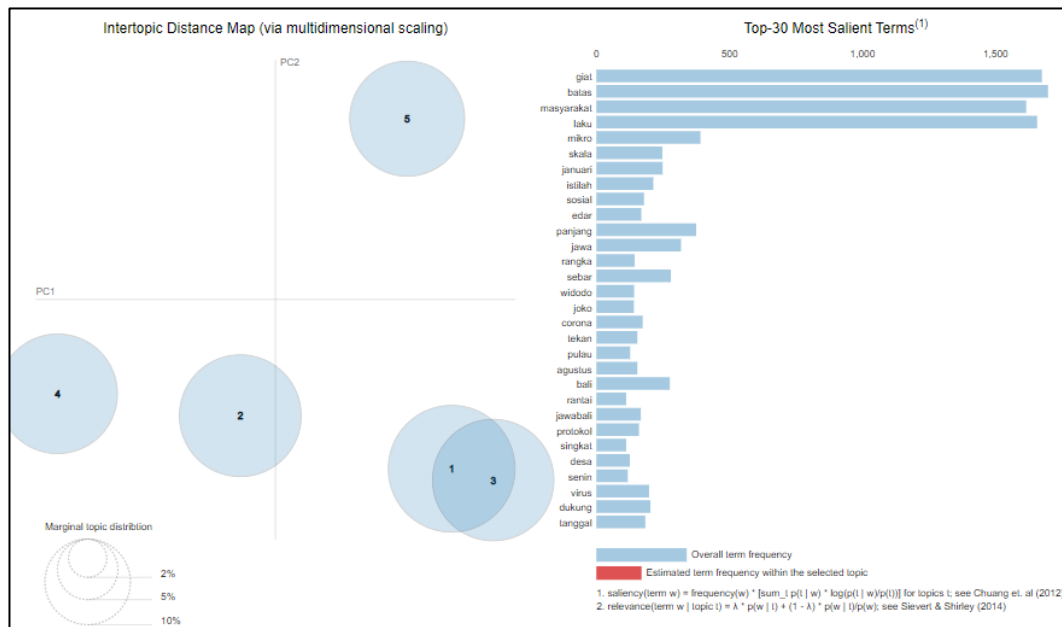
Terlihat dari grafik pada **Gambar 5.12** bahwa nilai koherensi meningkat secara signifikan mulai dari 1 topik menjadi 5 topik. Setelah mencapai 5 topik, nilai koherensi mengalami penurunan untuk 6 topik. Namun nilai koherensi yang diperoleh terus berfluktuasi hingga 10 topik dan mencapai skor tertinggi pada 5 topik. Hasil ini kemudian dijadikan acuan untuk analisis lebih lanjut, sehingga pembahasan mendalam akan terfokus pada 5 topik. Besarnya nilai koherensi berdasarkan grafik *coherence score* **Gambar 5.12** ditunjukkan pada **Tabel 5.18**

Tabel 5.18 Nilai Koherensi Topik PPKM dengan Unigram

| <i>Num Value</i> | <i>Coherence Value</i> |
|------------------|------------------------|
| 1 | 0.37439 |

| <i>Num Value</i> | <i>Coherence Value</i> |
|------------------|------------------------|
| 2 | 0.386547 |
| 3 | 0.433419 |
| 4 | 0.458055 |
| 5 | 0.507668 |
| 6 | 0.396852 |
| 7 | 0.445764 |
| 8 | 0.425812 |
| 9 | 0.459374 |
| 10 | 0.447163 |

Jumlah topik 5 memiliki peluang terbesar yaitu 0,507668 sehingga ini yang akan dijadikan acuan pemodelan topik. Kemudian dilakukan visualiasi untuk 5 topik yang terbentuk tersebut. Pemodelan topik akan sedikit sulit dipahami jika hanya dengan melihat kombinasi kata dan angka dari model yang terbentuk. Salah satu cara paling efektif untuk memahami data adalah melalui visualisasi *PyLDAvis*. Dengan begitu dapat memungkinkan untuk menginterpretasikan topik dalam model topik seperti diagram pada **Gambar 5.13**



Gambar 5.13 Visualisasi umum *Topic Modelling* PPKM dengan Unigram

Gambar 5.13 menunjukkan diagram terbagi menjadi 2 panel yaitu panel kiri dan panel kanan. Panel kiri dengan dimensi gelembung menggambarkan jarak antar topik dan area gelembung sesuai dengan prevalensi relatif topik di dalam korpus. Setiap gelembung mewakili sebuah topik dimana semakin besar gelembung, semakin tinggi persentase jumlah *tweet* dalam korpus tentang topik tersebut. Namun dilihat dari gelembung yang muncul pada 5 topik yang terbentuk menunjukkan bahwa area gelembung memiliki besaran yang cukup sama. Sehingga istilah frekuensi antara topik satu dengan lainnya tidak berbeda jauh bobotnya. Sedangkan panel kanan dijelaskan dengan menggunakan diagram batang sebagai banyaknya frekuensi kata atau istilah yang sering muncul. Dapat dilihat terdapat 30 istilah yang paling relevan untuk topik tertentu. Dari panel kanan hasil pemodelan topik diketahui diagram batang bahwa “oktober”, “polsek”, “nomor” adalah istilah yang sering dibahas publik ketika berbicara tentang PPKM. Dari 5 topik gelembung, terdapat topik yang beririsan yaitu topik 1 beririsan dengan topik 3. Kemudian memvisualisasikan istilah topik dengan *word cloud* memberikan gambaran kemungkinan adanya istilah dalam topik seperti pada **Gambar 5.14**.

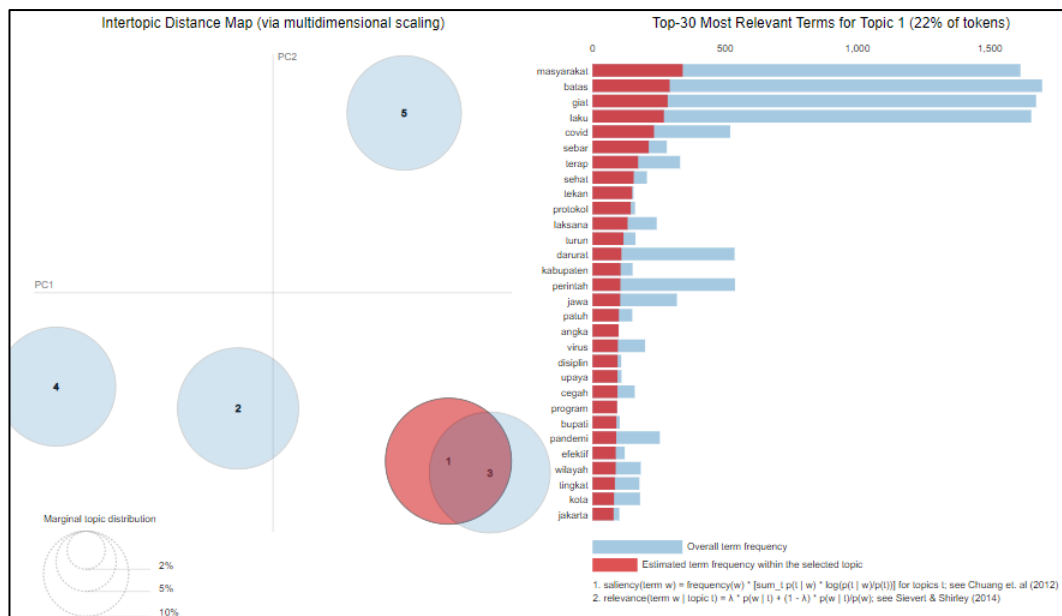


Gambar 5.14 *Word cloud* Topik 1 dengan Unigram

Word cloud disini merupakan kumpulan kata-kata yang digambarkan dalam berbagai ukuran. Visualisasi *word cloud* memberikan gambaran tentang kata-kata yang sering digunakan dalam komentar publik tentang PPKM. Semakin besar ukuran dari kata tersebut maka semakin sering kata tersebut dibicarakan. Dari *wordcloud* tersebut dapat dilihat kata-kata apa paling signifikan yang membangun segmen tertentu dari model yang sedang dibangun. **Gambar 5.14** menunjukkan *wordcloud* topik pertama tentang PPKM dimana “oktober”, “jawa”, “bali”,

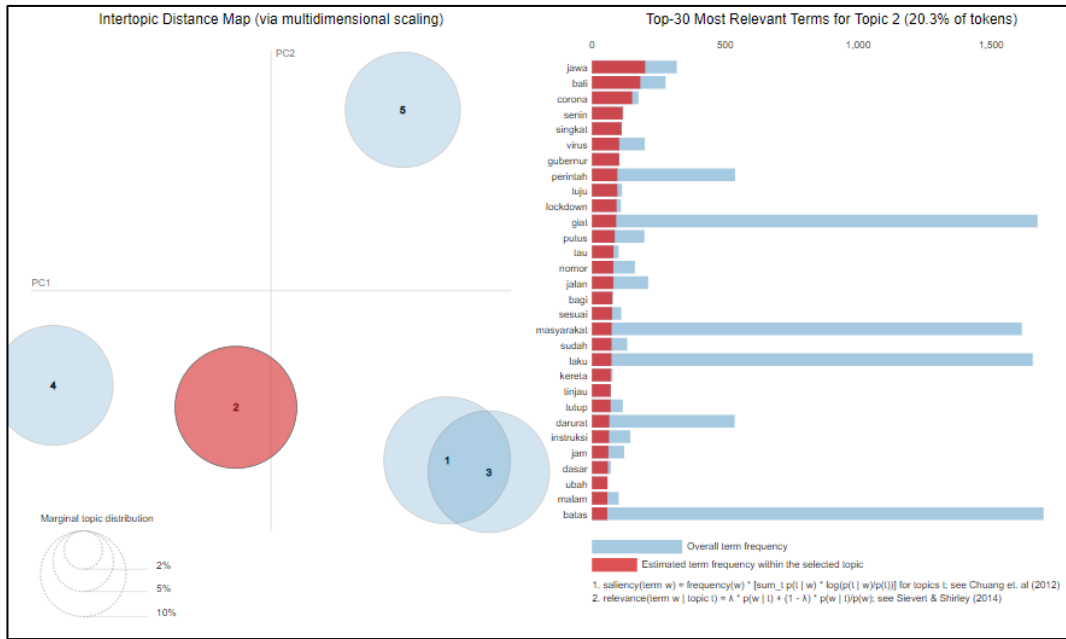
“panjang” merupakan kata yang banyak dibicarakan publik saat berbicara tentang PPKM. Selanjutnya untuk *wordcloud* yang lain terlampir di halaman 66.

Sebelumnya telah dilakukan visualisasi dengan *PyLDAvis* pada data *tweet* PPKM secara keseluruhan. Lalu untuk melihat frekuensi kata yang muncul dari masing-masing topik dapat dilakukan visualisasi pada topik 1 sampai topik 8 dimana setiap gelembung mewakili sebuah topik. Model topik yang baik akan memiliki gelembung besar dan tidak tumpang tindih yang tersebar di seluruh bagan. Seperti yang dapat dilihat dari grafik pada **Gambar 5.15** untuk hasil visualisasi topik 1.



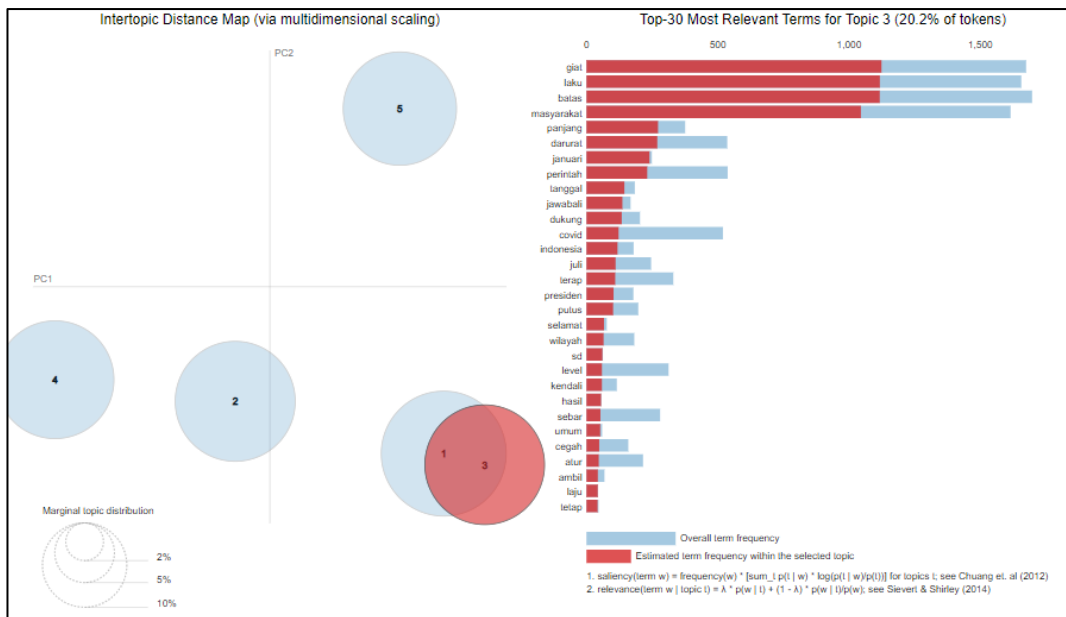
Gambar 5.15 Visualisasi Topik 1 dengan Unigram

Pada **Gambar 5.15** diagram batang berwarna merah memberikan perkiraan banyaknya istilah tertentu dihasilkan oleh topik tertentu. Seperti hasil diagram topik 1 ada sekitar 1500 lebih kata “masyarakat”, dan istilah ini digunakan sekitar 300 kali dalam topik 1. Kata dengan bilah merah terpanjang adalah kata yang paling banyak digunakan oleh *tweet* topik tersebut. Pada topik 1, menunjukkan bahwa kata diurutkan berdasarkan frekuensi terbanyak adalah masyarakat, batas, giat, laku, covid, sebar, terap, sehat, tekan, protokol, laksana, turun, darurat, kabupaten, perintah, jawa, patuh, angka, virus, disiplin, upaya, cegah, program, bupati, pandemi, efektif, wilayah, tingkat, kota, jakarta.



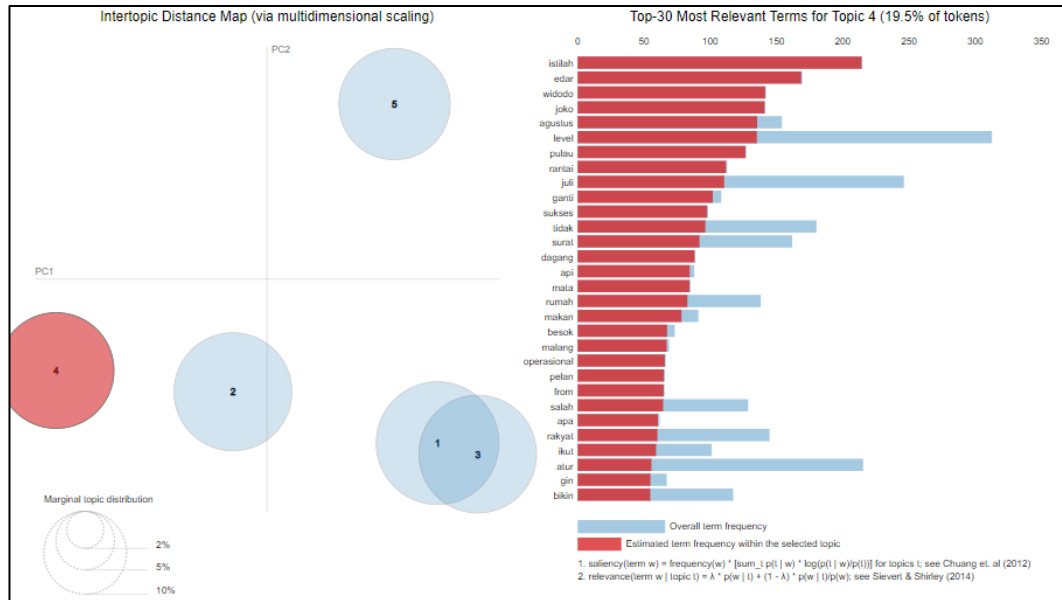
Gambar 5.16 Visualisasi Topik 2 dengan Unigram

Pada **Gambar 5.16** yang merupakan visualisasi topik 2, menunjukkan bahwa kata diurutkan berdasarkan kata frekuensi terbanyak adalah jawa, bali, corona, senin, singkat, virus, gubernur, perintah, tuju, lockdown, giat, putus, tau, nomor, jalan, bagi, sesuai, masyarakat, sudah, laku, kereta, tinjau, tutup, darurat, instruksi, jam, dasar, ubah, malam.



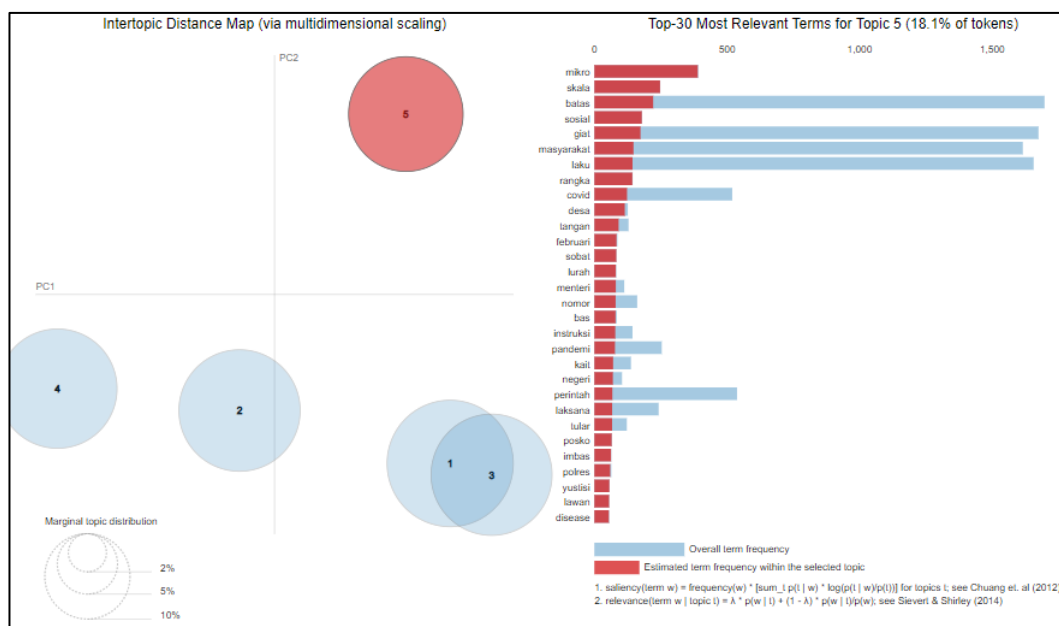
Gambar 5.17 Visualisasi Topik 3 dengan Unigram

Pada **Gambar 5.17** yang merupakan visualisasi topik 3, menunjukkan bahwa kata diurutkan berdasarkan kata frekuensi terbanyak adalah giat, laku, batas, masyarakat, panjang, darurat, januari, perintah, tanggal, jawa, bali, dukung, covid, indonesia, juli, terap, presiden, putus, selamat, wilayah, sd, level, kendali, hasil, sebar, umum, cegah, atur, ambil, laju, tetap.



Gambar 5.18 Visualisasi Topik 4 dengan Unigram

Pada **Gambar 5.18** yang merupakan visualisasi topik 4, menunjukkan bahwa kata diurutkan berdasarkan kata frekuensi terbanyak adalah istilah, edar, widodo, joko, agustus, level, pulau, rantai, juli, ganti, sukses, tidak, surat, dagang, api, mata, rumah, makan, besok, malang, operasional, pelan, from, salah, apa, rakyat, ikut, atur, gin, bikin.



Gambar 5.19 Visualisasi Topik 5 dengan Unigram

Pada **Gambar 5.19** yang merupakan visualisasi topik 5, menunjukkan bahwa kata diurutkan berdasarkan kata frekuensi terbanyak adalah mikro, skala, batas, sosial, giat, masyarakat, laku, rangka, covid, desa, tangan, februari, sobat, lurah, menteri, nomor, bas, instruksi, pandemi, kait, negeri, perintah, laksana, tular, posko, imbas, polres, yustisi, lawan, disease.

Kemudian setelah dilakukan visualisasi, dapat dilihat untuk hasil pemodelan topik dari 5 topik yang digunakan tentang perspektif publik tentang PPKM seperti **Tabel 5.19** berikut:

Tabel 5.19 Hasil Model Topik 5 dengan Unigram

| Topik | Model | Kesimpulan |
|-------|---|--------------------|
| 0 | 0.027*"masyarakat" + 0.023*"batas" + 0.022*"giat" + 0.021*"laku" + 0.018*"covid" + 0.017*"sebar" + 0.014*"terap" + 0.012*"sehat" + 0.012*"tekan" + 0.011*"protokol" | Protokol kesehatan |
| 1 | 0.017*"jawa" + 0.015*"bali" + 0.013*"corona" + 0.010*"senin" + 0.009*"singkat" + 0.009*"virus" + 0.009*"gubernur" | PPKM Jawa Bali |

| Topik | Model | Kesimpulan |
|-------|--|-------------------|
| | 0.008*"perintah" + 0.008*"tuju" + 0.008*"lockdown" | |
| 2 | 0.037*"mikro" + 0.024*"skala" + 0.021*"batas" + 0.017*"sosial" + 0.017*"giat" + 0.014*"masyarakat" + 0.014*"laku" + 0.014*"rangka" + 0.012*"covid" + 0.011*"desa" | PPKM mikro |
| 3 | 0.096*"giat" + 0.096*"laku" + 0.096*"batas" + 0.090*"masyarakat" + 0.023*"panjang" + 0.023*"darurat" + 0.021*"januari" + 0.020*"perintah" + 0.012*"tanggal" + 0.012*"jawabali" | Perpanjang PPKM |
| 4 | 0.019*"istilah" + 0.015*"edar" + 0.013*"widodo" + 0.012*"joko" + 0.012*"agustus" + 0.012*"level" + 0.011*"pulau" + 0.010*"rantai" + 0.010*"juli" + 0.009*"ganti" | Surat edaran PPKM |

Pada hasil model yang terbentuk dengan melihat visualisasi dengan *pyLDavis* terdapat beberapa model yang beririsan dimana pada model yang beririsan tersebut ada beberapa kata yang sama dibicarakan masyarakat terkait kebijakan PPKM yang diberlakukan pemerintah. Seperti disini untuk hasil model 1 dan model 3 beririsan yaitu keduanya kata yang sama banyak dibicarakan disini seperti kata laku, batas, giat, masyarakat, dan lainnya.

BAB VI PENUTUP

6.1. Kesimpulan

Berdasarkan hasil analisis LDA mengenai perspektif publik tentang PPKM pada media sosial *Twitter*, maka diperoleh beberapa kesimpulan sebagai berikut:

1. Diketahui gambaran umum data *tweet* tentang PPKM sebanyak 28530 baris, dengan diagram batang menunjukkan jumlah *tweet* tentang PPKM banyak dibicarakan pada bulan November dengan topik garis besar pembicaraan masyarakat adalah tentang penerapan PPKM dengan banyak level guna menekan angka kasus covid-19.
2. Hasil analisis sentimen didapatkan nilai persentase terbesar pada sentimen negatif yaitu sebesar 51,1%, yang artinya komentar tentang PPKM di *Twitter* kebanyakan masyarakat tidak mendukung ataupun menolak tentang adanya kebijakan ini.
3. *Topic modelling* dengan bigram trigram untuk koherensi topik dalam mengevaluasi jumlah topik yang paling sesuai, diperoleh 5 topik dengan nilai probabilitas adalah 0,496818. Hasil *topic modelling* dengan LDA tersebut adalah sebagai berikut :
 - a. Instruksi PPKM mikro.
 - b. Surat edar perpanjang PPKM.
 - c. Protokol kesehatan saat PPKM mikro.
 - d. PPKM Jawa Bali.
 - e. PPKM darurat.

Topic modelling dengan unigram untuk koherensi topik dalam mengevaluasi jumlah topik yang paling sesuai, diperoleh 5 topik dengan nilai probabilitas adalah 0,507668. Hasil *topic modelling* dengan LDA tersebut adalah sebagai berikut :

- a. Protokol kesehatan.
- b. PPKM Jawa Bali.
- c. PPKM mikro.
- d. Perpanjang PPKM.
- e. Surat edaran PPKM

6.2. Saran

Berdasarkan hasil dan pembahasan serta kesimpulan yang diperoleh, terdapat beberapa saran untuk pengembangan penelitian ke depan sebagai berikut:

1. Perlunya ahli bahasa dalam menginterpretasikan hasil topik.
2. Perlunya pengoptimalisasian saat *preprocessing* data mengingat masih banyak kata-kata singkatan yang memiliki makna sama namun berada pada entitas yang berbeda.

DAFTAR PUSTAKA

- A., V., & Sonawane, S. S. (2016). Sentiment Analysis of Twitter Data: A Survey of Techniques. *International Journal of Computer Applications*, 139(11), 5–15. <https://doi.org/10.5120/ijca2016908625>
- Ahmad Hania, A. (2017). Mengenal Artificial Intelligence, Machine Learning, & Deep Learning. *Jurnal Teknologi Indonesia*, June. <https://amt-it.com/mengenal-perbedaan-artificial-intelligence-machine-learning-deep-learning/>
- Apriani, A., Zakiyudin, H., & Marzuki, K. (2021). Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF System Penerimaan Mahasiswa Baru pada Kampus Swasta. *Jurnal Bumigora Information Technology (BITe)*, 3(1), 19–27. <https://doi.org/10.30812/bite.v3i1.1110>
- Audinovic, V., & Nugroho, R. S. (2021). Comics as Social Media Criticism of Community Activities Restrictions Enforcement Due to COVID-19 in Indonesia. 4(02), 295–313. <https://doi.org/10.37680/muharrir.v4i02.1041>
- Bin Muhammad Alkatiri, A., Nadiyah, Z., & Nasution, A. N. S. (2020). Opini Publik Terhadap Penerapan New Normal Di Media Sosial Twitter. *CoverAge: Journal of Strategic Communication*, 11(1), 19–26. <https://doi.org/10.35814/coverage.v11i1.1728>
- databoks. (2021). 10 Negara dengan Pengguna Twitter Terbanyak (per Juli 2021). <https://databoks.katadata.co.id/datapublish/2021/11/04/inilah-10-negara-dengan-pengguna-Twitter-terbanyak-ada-indonesia>
- Destarani, A. R., Slamet, I., & Subanti, S. (2019). Trend Topic Analysis using Latent Dirichlet Allocation (LDA) (Study Case: Denpasar People's Complaints Online Website). *Jurnal Ilmiah Teknik Elektro Komputer Dan Informatika*, 5(1), 50–58. <https://doi.org/10.26555/jiteki.v5i1.13088>
- Dewi, L. C., Meiliana, & Chandra, A. (2019). Social media web scraping using social media developers API and regex. *Procedia Computer Science*, 157, 444–449. <https://doi.org/10.1016/j.procs.2019.08.237>
- Dey, N., Ashour, A. S., & Nguyen, G. N. (2017). Deep learning for multimedia content analysis. *Mining Multimedia Documents*, 1(4), 193–203. <https://doi.org/10.1201/b21638>

- Fitri, V. A., Andreswari, R., & Hasibuan, M. A. (2019). Sentiment analysis of social media Twitter with case of Anti-LGBT campaign in Indonesia using Naïve Bayes, decision tree, and random forest algorithm. *Procedia Computer Science*, 161, 765–772. <https://doi.org/10.1016/j.procs.2019.11.181>
- Global Social Media Stats. (2021). <https://datareportal.com/social-media-users>
- Habibi, M., Priadana, A., Saputra, A. B., & Cahyo, P. W. (2021). Topic Modelling of Germas Related Content on Instagram Using Latent Dirichlet Allocation (LDA). 34(Ahms 2020), 260–264. <https://doi.org/10.2991/ahsr.k.210127.060>
- Hikmah, F. N., Basuki, S., & Azhar, Y. (2020). Deteksi Topik Tentang Tokoh Publik Politik Menggunakan Latent Dirichlet Allocation (LDA). *Jurnal Repositor*, 2(4), 415. <https://doi.org/10.22219/repositor.v2i4.52>
- Ilyas, F. (2021). Analisis Swot Kebijakan Pembatasan Sosial Berskala Besar (Psbb) Dan Pemberlakuan Pembatasan Kebijakan Masyarakat (Ppkm) Terhadap Dampak Ekonomi Di Tengah Upaya Menekan Laju Pandemi Covid-19. *Jurnal AKRAB JUARA Volume 6 Nomor 3 Edisi Agustus 2021 (190-198)*, 6, 190–198. <http://akrabjuara.com/index.php/akrabjuara/article/view/1559>
- Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. In *Multimedia Tools and Applications* (Vol. 78, Issue 11). <https://doi.org/10.1007/s11042-018-6894-4>
- Jongeling, R., Sarkar, P., Datta, S., & Serebrenik, A. (2017). On negative results when using sentiment analysis tools for software engineering research. *Empirical Software Engineering*, 22(5), 2543–2584. <https://doi.org/10.1007/s10664-016-9493-x>
- Kandukuri, M., & Haragopal, V. V. (2020). Textual Mining- Evaluation of Mann Ki Baat Repository. *Indian Journal of Science and Technology*, 13(27), 2711–2719. <https://doi.org/10.17485/ijst/v13i27.879>
- Kannan, S., Gurusamy, V., Vijayarani, S., Ilamathi, J. & Nithya, M. (2015). Preprocessing Techniques for Text Mining Preprocessing Techniques for Text Mining. *International Journal of Computer Science & Communication Networks*, 5(October 2014), 7–16.
- M. Yoserizal Saragih. (2021). The Role of Islamic Journalists in Consuring the

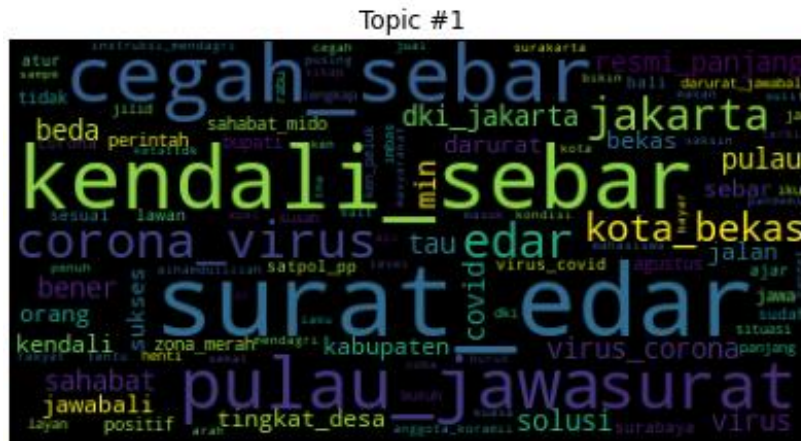
- Community towards the Provision of Hajj in 2021 in Covid-19 Pandemic. *Siasat*, 6(3), 137–145. <https://doi.org/10.33258/siasat.v6i3.98>
- Miharja, M., Salim, E., Nachrawi, G., Putranto, R. D., & Hendrawan, A. (2021). Implementation of Emergency Public Activity Restrictions (PPKM) in Accordance With Human Rights and Pancasila Principles. *BIRCI-Journal*, 15, 6855–6866. <https://doi.org/10.33258/birci.v4i3.2505>
- Morrison, S. C., Cuneo, R. C., Wainwright, D., & Stitz, R. W. (1993). Do the teaching hospitals of the University of Queensland really want a four-year medical course? - A guarded yes. *Medical Journal of Australia*, 159(5), 348–351. <https://doi.org/10.5694/j.1326-5377.1993.tb137878.x>
- Nisrina, F. D. (2020). *Implementasi Deteksi Topik Putusan Hakim dengan Latent Dirichlet Allocation (LDA)(Studi terhadap Putusan Tindak Pidana di Pengadilan Negeri Sleman, DI* <https://dspace.uui.ac.id/handle/123456789/23847>
- Novantirani, A., Sabariah, M. K., & Effendy, V. (2015). Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine. *E-Proceeding of Engineering*, 2(1), 1–7.
- Pfeffer, J., Mayer, K., & Morstatter, F. (2018). Tampering with Twitter’s Sample API. *EPJ Data Science*, 7(1). <https://doi.org/10.1140/epjds/s13688-018-0178-0>
- 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), 170–181.
- Rizal, M., Afrianti, R., & Abdurahman, I. (2021). Dampak Kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) bagi Pelaku Bisnis Coffe shop pada Masa Pandemi Terdampak COVID-19 di Kabupaten Purwakarta The Impact of the Policy for Implementing Community Activity Restrictions for Coffee Shop Busi. *Jurnal Inspirasi*, 12(1), 97–105.
- Rosid, M. A., Fitriani, A. S., Astutik, I. R. I., Mulloh, N. I., & Gozali, H. A. (2020). Improving Text Preprocessing for Student Complaint Document

- Classification Using Sastrawi. *IOP Conference Series: Materials Science and Engineering*, 874(1). <https://doi.org/10.1088/1757-899X/874/1/012017>
- Sah, S. (2020). Machine Learning: A Review of Learning Types. *ResearchGate*, July. <https://doi.org/10.20944/preprints202007.0230.v1>
- Setijohatmo, U. T., Rachmat, S., Susilawati, T., Rahman, Y., & Kunci, K. (2020). Analisis Metoda Latent Dirichlet Allocation untuk Klasifikasi Dokumen Laporan Tugas Akhir Berdasarkan Pemodelan Topik. *Prosiding The 11th Industrial Research Workshop and National Seminar*, 402–408.
- Taofik Krisdiyanto, E. M. O. N. (2021). Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan Naïve Bayes Clasifiers. *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer Dan Teknologi Informasi*, 7(1), 32–37. <http://ejournal.uin-suska.ac.id/index.php/coreit/article/view/12945>
- Wati, R., Ernawati, S., Studi, P., Informasi, S., Bina, U., Informatika, S., Studi, P., Informasi, S., Nusa, U., & Timur, J. (2021). Analisis Sentimen Persepsi Publik Mengenai PPKM Pada Twitter Berbasis SVM Menggunakan Python. 06, 240–247.
- Wehle, H. (2017). *ML – AI- COGNITIVE*. August.
- Yang, S., & Zhang, H. (2018). Text Mining of Twitter Data Using a Latent Dirichlet Allocation Topic Model and Sentiment Analysis Text Mining of Twitter Data Using a Latent Dirichlet Allocation Topic Model and Sentiment Analysis View project. *International Journal of Computer and Information Engineering*, 12(7), 525–529. <https://www.researchgate.net/publication/335106801>

LAMPIRAN

| Lampiran | Link |
|--|---|
| Lampiran 1 Data <i>Tweet</i> PPKM | https://github.com/desyendriani/Data-Tweet-PPKM |
| Lampiran 2 <i>Script Scapping</i> | https://github.com/desyendriani/Script-Scapping/blob/main/Scapping%20Data%20(python).txt |
| Lampiran 3 <i>Script Grafik</i> | https://github.com/desyendriani/Grafik/blob/main/Grafik%20Visualisasi.txt |
| Lampiran 4 <i>Script Analisis Sentimen</i> | https://github.com/desyendriani/Analisis-Sentimen/blob/main/Analisis%20Sentimen.txt |
| Lampiran 5 <i>Script TF-IDF</i> | https://github.com/desyendriani/TF-IDF/blob/main/TF-IDF%20(Python).txt |
| Lampiran 6 <i>Script Topic Modelling</i> | https://github.com/desyendriani/Topic-Modelling-LDA |

Lampiran 7 *Word cloud* Topik 2 PPKM dengan Bigram Trigram



Lampiran 8 Word cloud Topik 3 PPKM dengan Bigram Trigram

Topic #2



Lampiran 9 Word cloud Topik 4 PPKM dengan Bigram Trigram

Topic #3

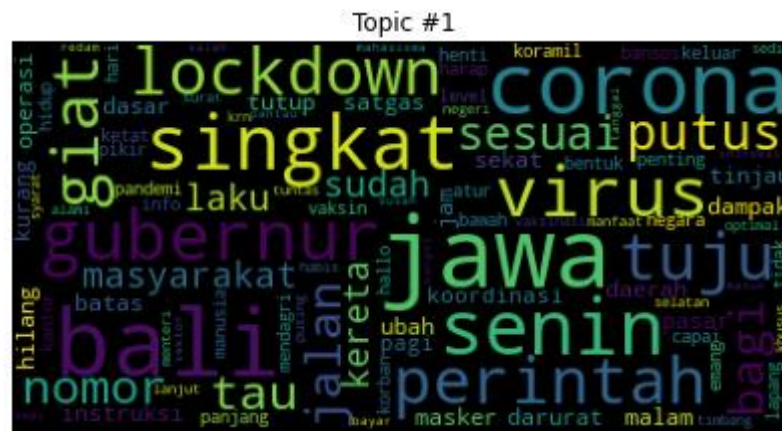


Lampiran 10 Word cloud Topik 5 PPKM dengan Bigram Trigram

Topic #4



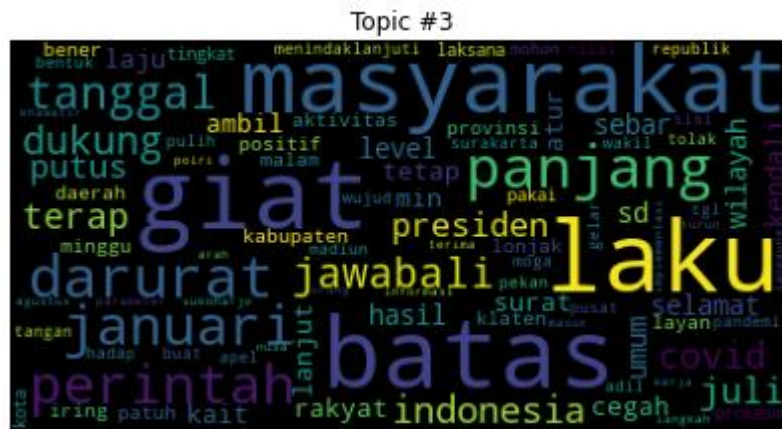
Lampiran 11 Word cloud Topik 2 PPKM dengan Unigram



Lampiran 12 Word cloud Topik 3 PPKM dengan Unigram



Lampiran 13 Word cloud Topik 4 PPKM dengan Unigram



Lampiran 14 Word cloud Topik 4 PPKM dengan Unigram

Topic #4

