TOPIC MODELLING PADA BERITA *ONLINE* UNTUK PENCARIAN PENYEBAB RESESI DENGAN METODE LDA

SKRIPSI

Oleh:
TENGKU SURYA AL FURQAN
NIM.19650094

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2023

TOPIC MODELLING PADA BERITA *ONLINE* UNTUK PENCARIAN PENYEBAB RESESI DENGAN METODE LDA

SKRIPSI

Diajukan Kepada:
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan Dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh:
TENGKU SURYA AL FURQAN
NIM. 19650094

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG
202

HALAMAN PERSETUJUAN

***TOPIC MODELLING* PADA BERITA *ONLINE* UNTUK PENCARIAN PENYEBAB RESESI DENGAN METODE LDA**

SKRIPSI

Oleh:
TENGKU SURYA AL FURQAN
NIM. 19650094

Telah Diperiksa dan Disetujui untuk Diuji
Tanggal: 13 Desember 2023

Pembimbing I	Pembimbing II
Dr. Fachrul Kurniawan, M.MT, IPM	Dr. Fresy Nugroho, M.T
NIP. 19771020 200912 1 001	NIP. 19710722 201101 1 001

Mengetahui,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Dr. Fachrul Kurniawan, M.MT
NIP. 19771020 200912 1 001

iii

HALAMAN PENGESAHAN

TOPIC MODELLING PADA BERITA *ONLINE* UNTUK PENCARIAN PENYEBAB RESESI DENGAN METODE LDA

SKRIPSI

Oleh:
TENGKU SURYA AL FURQAN
NIM. 19650094

Telah Dipertahankan Di Depan Dewan Penguji Skripsi
Dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal: 13 Desember 2023

Susunan Dewan Penguji

Ketua Penguji	: Dr. Muhammad Faisal, M.T NIP. 19740510 200501 1 007	( )
Anggota Penguji I	: Puspa Miladin Nuraida Safitri A Basid, M.Kom NIP. 19930828 201903 2 018	
Anggota Penguji II	: Dr. Fachrul Kurniawan, M.MT, IPM NIP. 19771020 200912 1 001	
Anggota Penguji III	: Dr. Fresy Nugroho, M.T NIP. 19710722 201101 1 001	( )

Mengetahui dan Mengesahkan,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Dr. Fachrul Kurniawan, M.MT
NIP. 19771020 200912 1 001

iv

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama	:	Tengku Surya Al Furqan
NIM	:	19650094
Fakultas / Jurusan	:	Sains dan Teknologi / Teknik Informatika
Judul Skirpsi	:	*Topic Modelling* Pada Berita *Online* Untuk Pencarian Penyebab Resesi Dengan Metode LDA

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 13 Desember 2023
Yang membuat pernyataan,

Tengku Surya Al Furqan
NIM. 19650094

HALAMAN MOTTO

“Jangan Terlalu Banyak Berpikir, Kerjakan Saja”

HALAMAN PERSEMBAHAN

Puja dan puji syukur atas kehadirat Allah subhanahu wa ta'ala, serta shalawat dan salam bagi Rasul-Nya Penulis mepersembahkan hasil karya ini kepada:

Orang tua penulis yang saya sayangi dan cintai, Bapak Dang Marta dan Ibu Evi Yulianti, yang terus mendukung dan menyemangati dan tak lupa terus mendo'akan penulis.

Para dosen pembimbing penulis, Bapak Fachrul Kurniawan, M.MT, IPM dan Bapak Fresy Nugroho, M.T yang senantiasa telah memberi saran dan kemudahan dalam membantu penulis untuk menyusun karya tulis ini.

Seluruh dosen dan jajaran civitas akademik Jurusan Teknik Informatika Universitas Islam Negeri Maulana Malik Ibahim Malang yang telah memberi pelajaran dan ilmu dan keahlian dalam menyelesaikan karya tulis ini.

Seluruh teman-teman Alliance of Informatics Engineering (ALIEN) angakatan 19 yang turut serta saling membantu, memberi semangat, dan saling memberi dukungan kepada penulis untuk menyelesaikan karya tulis ini. Serta teman – teman SMA yang terus terjalin hubungannya dan ikut memberi bantuan dan mendukung secara mental agar penulis dapat menyelesaikan karya tulis ini.

KATA PENGANTAR

Segala puji dan syukur senantiasa penulis sampaikan pada Allah SWT yang berkat rahmat serta hidayat-nya, sehingga penulis dapat menyempurnakan skripsi ini pada waktunya. Sholawat serta salam tercurahkan pada Nabi Muhammad SAW yang telah menuntun umat manusia ke jalan yang lebih baik.

Penulis mengucapkan rasa terima kasih yang begitu besar kepada seluruh pihak yang memberikan dukungan dan membantu sempurnanya skripsi ini. Ucapan terimakasih penulis ditujukan kepada:

1. Prof. Dr. M. Zainuddin, M.A., selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Prof. Dr. Hj. Sri Harini, M.Si., selaku dekan Fakultas Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Fachrul Kurniawan, M.MT, selaku ketua Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Kedua orang tercinta yang selalu memberi dukungan dan do'a pada penulis hingga rampungnya skripsi ini.
5. Dr. Fachrul Kurniawan, M.MT, dan Dr. Fresy Nugroho, M.T, selaku dosen pembimbing I dan II yang telah membimbing dan memberikan arahan kepada peneliti sehingga dapat menyelesaikan skripsi ini.
6. Dr. Muhammad Faisal, M.T, dan Puspa Miladin Nuraida Safitri A Basid, M.Kom selaku dosen penguji I dan II yang telah memberikan, kritik serta saran kepada penulis hingga ujian skripsi dengan penuh kesabaran.

7. Kawan-kawan Alliance of Informatics Engineering (ALIEN) Angakatan 2019, khususnya grup Team Sunmori dan Robi'atul Adawiyah yang senantiasa selalu memberikan semangat dan dukungan dalam berjuang bersama dalam mengejar gelar S.Kom dan pengalaman di universitas yang sama.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih jauh dari kata sempurna. Maka dari itu penulis menerima saran, kritik dan masukan yang bersifat membangun sehingga dapat menjadi lebih baik kedepannya. Penulis berharap semoga skripsi ini dapat memberikan manfaat untuk kedepannya. Wassalamu'alaikum warahmatullahi wabarakatuh.

Malang, 13 Desember 2023

Penulis

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

ABSTRAK

Surya Al-Furqan, Tengku.2023. ***Topic Modelling*** **Pada Berita** ***Online*** **Untuk Pencarian Penyebab Resesi Dengan Metode LDA.** Skripsi Jurusan Teknik Informatika, Fakultas sains dan Teknologi. Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing (I) Dr. Fachrul Kurniawan, M.MT, IPM, (II) Dr. Fresy Nugroho, M.T.

Kata Kunci: *Topic Modelling, Latent Dirichlect Allocation,* berita online, resesi

Berita online telah menjadi teknologi informasi yang digunakan bebagai orang untuk memperoleh informasi utama dalam era digital ini Penggunaan berita online sebagai penyedia informasi semakin berkembang semasa pandemi. Setelah periode pandemi, banyak berita online yang menyampaikan permasalahan mengenai perekonomian yang akan tertimpa resesi. Peningkatan jumlah berita yang terkait kondisi ekonomi resesi setelah pandemi tanpa diketahui bahasannya hal ini dapat memperkeruh situasi dengan menyebarkan kebingungan, kepanikan, atau informasi yang menyesatkan kepada masyarakat. Penelitian ini dilakukan untuk membangun model dan mengidentifikasi berbagai topik yang menyusun artikel mengenai resesi di Indonesia menggunakan metode *latent dirichlect allocation*. Data yang digunakan merupakan proses *crawling* pada beberapa berita online Indonesia yang diproses untuk pembersihan dengan *text-processing*. Data yang telah terproses lalu di bangun korpus lalu dibangun model LDA yang menghasilkan model yang digunakan pada iterasi 1000 dengan nilai *Cv* sebesar 0,4473 dan *Umass* sebesar -3,9757. Hasil dari model menghasilkan 20 topik dengan topik ke-9 membahas penyebab resesi dengan peringkat tertinggi dengan nilai 100% dan topik-4, dan 16 membahas penyeab resesi dengan peringkat terendah dengan nilai 61,54%.

ABSTRACT

Al-Furqan, Tengku Surya. 2023. **Topic Modeling in Online News to Find the Causes of Recessions Using the LDA Method**. Thesis, Informatics Engineering Department, Faculty of Science and Technology. State Maulana Malik Ibrahim Islamic University Malang. Supervisors (I) Dr. Fachrul Kurniawan, M.MT, IPM, (II) Dr. Fresy Nugroho, M.T.

Online news has become an information technology used by various people to obtain key information in this digital era The use of online news as an information provider has grown during the pandemic. After the pandemic period, a lot of online news conveyed problems about the economy that would be hit by a recession. The increase in the number of news related to recessionary economic conditions after the pandemic without knowing the discussion can worsen the situation by spreading confusion, panic, or misleading information to the public. This research was conducted to build a model and identify various topics that make up articles about the recession in Indonesia using the latent direct allocation method. The data used is a crawling process on several Indonesian online news which is processed for cleaning with text-processing. The data that has been processed then built a corpus and then built an LDA model that produces a model that is used at iteration 1000 with a Cv value of 0.4473 and Umass of -3.9757. The results of the model produced 20 topics with the 9th topic discussing the cause of the recession with the highest rank with a value of 100% and topic-4, and 16 discussing the cause of the recession with the lowest rank with a value of 61.54%.

Keywords: *Topic Modelling, Latent Dirichlect Allocation, online news, recession*

الملخص

.الفرقان، تينكو سوريا .٢٠٢٣ .نمذجة المواضيع في الأخبار عبر الإنترنت للعثور على أسباب الركود باستخدام طريقة لدا .أُطرُوحَة قسم الهندسة المعلوماتية، كلية العلوم والتكنولوجيا .جامعة مولانا مالك إبراهيم الإسلامية مالانج. المشرفون (١)د .فخرول كورنياوان (٢)د .فريزي نوجروهو

أصبحت الأخبار عبر الإنترنت بمثابة تكنولوجيا معلومات يستخدمها أشخاص مختلفون للحصول على معلومات أساسية في هذا العصر الرقمي، وقد نما استخدام الأخبار عبر الإنترنت كمزود للمعلومات خلال الوباء .بعد فترة الوباء، نقلت الكثير من الأخبار عبر الإنترنت مشاكل حول الاقتصاد الذي سيتضرر من الركود .إن زيادة عدد الأخبار المتعلقة بالأوضاع الاقتصادية الركود بعد الوباء دون معرفة المناقشة يمكن أن يؤدي إلى تفاقم الوضع من خلال نشر البلبلة أو الذعر أو المعلومات المضللة للجمهور .تم إجراء هذا البحث لبناء نموذج وتحديد الموضوعات المختلفة التي تشكل مقالات حول الركود في إندونيسيا باستخدام طريقة التخصيص المباشر الكامن .البيانات المستخدمة عبارة عن عملية زحف على العديد من الأخبار الإندونيسية عبر الإنترنت والتي تتم معالجتها للتنظيف باستخدام معالجة النصوص. البيانات التي تمت معالجتها ثم قامت ببناء مجموعة ثم بناء نموذج لدا الذي ينتج نموذجًا يتم استخدامه بقيمة ٣,٩٧٥٧.. أنتجت نتائج النموذج ٢٠ موضوعا، الموضوع التاسع Umass تبلغ ٠,٤٤٧٣ و Cv في التكرار ١٠٠٠ بقيمة يناقش سبب الركود ذو الرتبة الأعلى بقيمة ١٠٠٪ والموضوع-٤، و١٦ يناقش سبب الركود ذو الرتبة الأقل بقيمة ٦١,٥٤

الكلمات المفتاحية :نمذجة الموضوع، التخصيص المباشر الكامن، الأخبار عبر الإنترنت، الركود

BAB I

PENDAHULUAN

1.1 Latar Belakang

Semakin tingginya laju pertumbuhan teknologi pada zaman modern ini menghasilkan berbagai macam teknologi di sekitar kita. Pada berbagai bidang telah terjadi pertumbuhan teknologi. bidang analisa data adalah salah satunya. Data dapat ditemukan dimana-mana. Dengan adanya internet, data dapat diperoleh dan dianalisa. Ribuan *bit* data masuk ke dalam jaringan komputer kita, *web*, *database*, *server* dan berbagai alat untuk menyimpan data digunakan setiap hari terutama dari bisnis, hubungan sosial, ilmu pengetahuan dan teknik, kesehatan, dan semua aktivitas sehari-hari. Meningkatnya jumlah dari volume data yang dapat di peroleh ini merupakan bukti dari komputerisasi masyarakat dan perkembangan cepat *tools* untuk mengumpulkan dan penyimpanan data yang kuat (Niu et al., 2021).

Dengan kemajuan teknologi informasi, informasi sudah tersebar di berbagai domain seperti kesehatan, astronomi, *web sosial*, dan *geoscience* (Ghani et al., 2019). Konten sosial media seperti *tweet*, *comment*, *post*, dan *review* dapat menjadi kontributor besar dalam pembuatan *big data* baik dari *platform* tertentu atau *website* yang berbeda (Martínez-Castaño et al., 2020).

Salah satu pengambilan *big data* yang sangat mudah ditemui adalah sebuah artikel atau berita yang diterbikan secara *online*. Terdapat berbagai macam *website* berita yang memberikan kebebasan dalam menulis sebuah artikel. Artikel tersebut dapat dibuat oleh siapa saja dan dimana saja. Artikel ini dapat diakses

dengan mudah oleh siapapun karena semakin berkembangnya kemajuan teknologi informasi.

Pentingnya keberadaan *big data* dari berbagai artikel telah mendatangkan berbagai macam hal baru terutama dalam bidang analisis dan kecerdasan buatan. Analisis media informasi menggunakan berbagai macam alogaritma mulai dari yang tradisional hingga alogaritma yang mengimplementasikan *machine learning* di dalamnya sebagai bagian dari penelitian. Sebagai contoh, produk sebuah toko dapat melakukan *forecasting* atau prediksi produk yang akan laku dijual pada tokonya. Atau perusahaan dapat mengelompokkan apakah investasi akan mendapatkan keuntungan berapa persen berdasarkan kelompok penjualannya (Maulana & Fajrin, 2018).

Setelah masa pandemi ini banyak artikel mengenai menurunnya perekonomian dan semakin dekatnya negara – negara besar di dunia dalam ambang resesi. Resesi adalah penurunan atau kemerosotan yang signifikan dalam aktivitas ekonomi yang menyebar ke semua bidang ekonomi, termasuk produksi, kesempatan kerja penuh, pendapatan riil, dan lainnya, yang biasanya terjadi dalam beberapa bulan jangka pendek (Jacob & Waibot, 2022). Berdasarkan definisi tersebut, kejadian yang tergolong sebagai resesi dapat terjadi ketika kondisi suatu negara mengalami penurunan nilai pertumbuhan ekonomi, fluktuasi, inflasi, kesenjangan *Gross National Product* (GNP), produk domestik bruto negatif selama dua kuartal berturut-turut, ini menunjukkan bahwa aktivitas ekonomi seperti produksi, distribusi, konsumsi, investasi, dan sebagainya akan mengalami penurunan, yang memiliki dampak negatif pada banyak hal, termasuk PHK, Selain

itu, krisis yang terjadi di suatu negara atau bahkan di seluruh dunia juga menjadi salah satu factor penyebab resesi (Asrirawan et al., 2022; Blandina et al., 2020; Gumelar et al., 2023; Jacob & Waibot, 2022).

Berita ini disampaikan di berbagai media seperti twitter, youtube, dan artikel berita *online*. Di mana tertulis bahwa negara – negara global akan terkena resesi dimana akan berdampak juga pada ekonomi Indonesia yang berimbas bisa mengalami resesi juga. Hal ini dikaitkan dengan adanya permasalahan yang sedang dan telah terjadi di dunia saat ini yaitu pemulihan pasca Covid – 19 dan perang antara Rusia dan Ukraina (Mahdiyan, 2023).

Banyak dari artikel menjelaskan Indonesia akan terkena dampak dari perang Rusia dengan Ukraina karena hubungan antara perang yang mengakibatkan negara – negara dengan negara G7 mengalami penurunan dalam beberapa dekade terakhir. Dari laporan yang diberikan oleh Bank Dunia memperkirakan harga energi akan meningkat lebih dari 50% pada tahun 2022. Peningkatan ini terjadi terutama di negara Uni Eropa akibat konflik semakin memburuk (Zakeri et al., 2022).

Pada artikel yang ditulis pada berita di Indonesia, banyak artikel mengenai resesi yang tidak menjelaskan secara detail penyebab dan asal – usul mengenai pemicu adanya resesi di Indonesia. Berita tersebut tidak disampaikan secaranya menyeluruh sehingga terhambatnya informasi yang di dapat juga mengakibatkan turunnya minat pembaca untuk melakukan analisis dan memanfaatkan artikel tersebut untuk mengetahui hubungan antara topik satu dengan topik lainnya. Untuk mengatasi hal – hal tersebut dapat menggunakan teknik *topic modelling*. Selain itu, artikel mengenai resesi juga dikaitkan dengan hal – hal yang tidak berhubungan

dengan ekonomi sehingga banyaknya artikel yang tidak sesuai dengan topiknya dapat menghambat informasi yang disampaikan.

Terdapat beberapa penelitian yang telah membahas hal tersebut, seperti (Naury et al., 2021) yang menggunakan *topic modelling* untuk pemodelan topik terhadap hasil sentimen berita tajuk Indonesia. Pada penelitian (Alfanzar et al., 2020) menggunakan untuk melihat tren dari ketertarikan pada program studi Sastra Inggris di universitas Islam Negeri Sunan Ampel. Pada penlitian (Firdaus et al., 2020) digunakan *topic modelling* untuk mengetahui kepuasan pelanggan terhadap aplikasi didik ruangguru. Penelitian lainnya dilakukan oleh (Al-khairi et al., 2017) dalam mendeteksi topik trendi di twitter menggunakan distribusi laten dirichlet allocation, jumlah topik yang optimal adalah 20 Topik.

Dengan adanya *topic modelling* diharapkan bahwa penulis dapat menjelaskan topik dan mengetahui hal-hal yang menjadi penyebab resesi ekonomi. Seperti yang dijelaskan dalam Q.S An – Nur ayat 15 yang berbunyi :

إِذْ تَلَقَّوْنَهُۥ بِأَلْسِنَتِكُمْ وَتَقُولُونَ بِأَفْوَاهِكُم مَّا لَيْسَ لَكُم بِهِۦ عِلْمٌ وَتَحْسَبُونَهُۥ هَيِّنًا وَهُوَ عِندَ ٱللَّهِ عَظِيمٌ

"(Ingatlah) ketika kamu menerima (berita bohong) itu dari mulut ke mulut; kamu mengatakan dengan mulutmu apa yang tidak kamu ketahui sedikit pun; dan kamu menganggapnya remeh, padahal dalam pandangan Allah itu masalah besar." (QS An-Nur: 15).

Berdasarkan tafsir Al – Wajiz dijelaskan arti ayat diatas sebagai berikut: Ingatlah di waktu kamu menerima berita bohong itu dari mulut ke mulut dan kamu katakan dengan mulutmu apa yang tidak kamu ketahui sedikit pun. Lalu kamu menganggapnya suatu yang ringan saja. Padahal kesalahan itu pada sisi Allah

adalah besar. Sedangkan pada tafsir Al – Azhar dijelaskan ayat itu berkaitan dengan ayat sebelumnya tentang tidak boleh menajatuhkan tuduhan-tuduhan belaka. Dimana di mata Allah mereka adalah pembohong belaka. Tetapi di sisi munafik, kebohongan itulah yang menjadi benar dan nyata. Selanjutnya pada an - nur ayat 14 dijelaskan akan ada azab besar nantinya yang akan ditimpakan tuhan karena menyebarkan berita bohong dan dilanjutkan ayat pada an – nur ayat 15 dimana berita itu langsung disambut dengan lidah, tidak tahu dari mana pangkalnya dan apa ujungnya. Dari lidah ke lidah dan membuat orang – orang panik.

Dari sini dapat dilihat bahwa pentingnya untuk mencari tahu informasi terlebih dahulu sebelum mennyampaikannya ke publik agar tidak menjadi suatu kebohongan. Pentingnya untuk menyikapi dan mengetahui isi atau topik dari suatu informasi terlebih dahulu sebelum menyebarkannya agar tidak mengakibatkan keributan Ketika disebarkan.

Rasulullah *shallallahu ‘alaihi wa sallam* dengan tegas mengatakan,

كَفَى بِالْمَرْءِ كَذِبًا أَنْ يُحَدِّثَ بِكُلِّ مَا سَمِعَ

“Cukuplah seseorang dikatakan sebagai pendusta apabila dia mengatakan semua yang didengar.” (HR. Muslim no.7)

Jika tidak ada manfaatnya atau bahkan justru berpotensi menimbulkan salah paham, keresahan atau kekacauan di tengah-tengah masyarakat dan hal-hal yang tidak diinginkan lainnya, maka hendaknya tidak langsung disebarkan (diam) atau minimal menunggu waktu dan kondisi dan tepat. Rasulullah *shallallahu ‘alaihi wa sallam* bersabda,

وَمَنْ كَانَ يُؤْمِنُ بِاللَّهِ وَالْيَوْمِ الآخِرِ فَلْيَقُلْ خَيْرًا أَوْ لِيَصْمُتْ

"Barangsiapa beriman kepada Allah dan hari akhir, hendaklah berkata yang baik atau diam." (HR. Bukhari no. 6018 dan Muslim no. 74)

Penggunaan *data mining* dan machine learning dapat dimanfaatkan dalam mencari tahu lebih dalam mengenai hal – hal yang berkaitan dengan resesi. Suatu cara atau Langkah-langkah untuk mendapatkan relasi yang tepat, pola, dan kecenderungan dengan memeriksa suatu data dalam kumpulan yang disimpan di dalam suatu penyimpanan dengan memanfaatkan suatu cara untuk mengenal pola seperti teknik pada statistik dan matematika merupakan pengertian dari Data Mining. (Nastuti, 2019). Selanjutnya menggunakan unsupervised learning dengan metode *topic modelling* yaitu *Latent Dirichlet Allocation* (LDA). LDA dipilih karena banyak digunakan dalam pencarian topik pada Kumpulan dokumen dan juga digunakan untuk menganalisis hubungan antar dokumen dalam korpus (Qomariyah et al., 2019). Penelitian ini nantinya untuk mengetahui topik atau tajuk yang sering digunakan untuk menentukan terjadinya resesi di artikel Indonesia.

1.2 Pernyataan Masalah

Seperti penjelasan yang telah diberikan oleh penulis sebelumnya, Pernyataan masalah penelitian ini adalah bagaimana melakukan pemodelan topik pada kumpulan artikel mengenai resesi pada artikel berita *online* Indonesia dan ditemukan relevansi dan korelasi antara topik dan parameter terduga menjadi penyebab resesi di Indonesia.

1.3 Tujuan Penelitian

Penelitian ini bertujuan untuk membangun model dan mengidentifikasi berbagai topik yang menyusun artikel mengenai resesi di Indonesia.

1.4 Hipotesis

Dengan menggunakan metode *Latent Dirichlet Allocation* dapat ditemukan relevansi dan korelasi antara topik dan parameter terduga menjadi penyebab resesi di Indonesia.

1.5 Manfaat Penelitian

1. Dapat mengetahui hal – hal yang menjadi penyebab dan bahasan pada resesi di Indonesia.
2. Dapat membantu memprediksi adanya resesi di Indonesia tahun 2023.
3. Dengan mengetahui hasil prediksi ini, diharapkan bisa membantu dalam mempersiapkan dalam menghadapi prediksi ekonomi.
4. Dapat mengetahui pemanfaatan *data mining* dalam mengumpulkan data untuk prediksi.

1.6 Batasan Masalah

1. Data yang digunakan adalah data *text mining* dari artikel berita yaitu CNN Indonesia, Detik.com, Kompas.com, Antaranews, dan Idntimes dari tahun 2022 - 2023
2. Obyek di dalam penelitian ini adalah berfokus terhadap resesi dan topik yang mendekatinya.

3. Metode *topic modelling* yang digunakan adalah *Latent Dirichlet Allocation* (LDA).
4. Tools yang digunakan pada penilitian ini adalah web scrapper dan google colab
5. Bahasa pemrograman yang digunakan pada penelitian ini menggunakan Bahasa pemrograman python

BAB II

STUDI PUSTAKA

2.1 Penelitian Terkait

Perancangan suatu penelitian diperlukan hal utama yang membuat penelitian dapat diakui. Hal itu dilakukan dengan adanya dukungan dari penelitian sebelumnya yang telah dilakukan dan sudah ada. Selain itu juga berhubungan dengan penelitian tersebut.

Penelitian yang telah dilakukan oleh Al-khairi (2017) tentang deteksi topik fashion pada twitter menggunakan *Latent Dirichlet Allocation* mengambil data yang bersumber dari twitter dengan topik terfokus pada fashion dari bulan Agustus 2017 – September 2017. Data kemudian di praprocessing dengan *gibss sampling* yang selanjutnya dilakukan proses olah kata umumnya yaitu *lowercasing*, *stopword*, *steaming*, dan TF-IDF. Selanjutnya menggunakan metode LDA untuk mengetahui topik yang sedang dibahas dengan hasil evaluasi yaitu *topic coherence* PMI terbaik dengan score 6.272 (Al-khairi et al., 2017). pada penelitian yang di usung, akan digunakan hasil evaluasi dengan menggunakan *Coherence score*.

Penelitian yang telah dilakukan oleh Chairullah (2020) tentang *topic modelling* pada sentiment terhadap headline berita *online* berbahasa Indonesia menggunakan LDA dan LSTM. Penelitian ini menggunakan dataset dari webscrapping dari berbagai berita *online* Indonesia dengan topik terfokus pada covid 19 atau *coronavirus disease* 2019. Data yang telah dikumpukan selanjutnya dilabeli positif, negatif, dan netral. LSTM digunakan untuk klasifikasi teks berdasarkan sentimennya yang selanjutnya menggunakan LDA untuk mengetahui

topik pembahasan setiap sentiment yang telah dibagi. Penelitian menghasilkan menghasilkan model dengan nilai akurasi tertinggi sebesar 71.13% dan akurasi terendah sebesar 63.48%.(Chairullah, 2020) Metode pada penelitian ini akan digunakan sebagai acuan pada penelitian.

Nurlayli dan Nasichudding (2019) menggunakan *topic modelling* untuk mengetahui topik yang sering digunakan oleh penelitian dosen JPTEI Universitas Negeri Yogyakarta. Data yang diambil adalah kumpulan judul dari penelitian dosen UNY yang bersumber dari google schoolar. Untuk prosesnya menggunakan *tokenizing*, *stopword*, dan *bigram* untuk pembagian 2 kata setiap kalimat. Lalu menggunakan metode LDA untuk mengetahui topik yang digunakan pada penelitian. Penelitian menghasilkan topik 0 membahas tentang Pendidikan Vokasi, topik 1 membahas tentang pengembangan Sistem, topic 2 mebahas tentang interpretasikan berbicara tentang media pembelajaran, topik 3 membahas tentang Sistem pembelajaran di SMK (Nurlayli & Nasichuddin, 2019). Metode pada penelitian ini akan digunakan sebagai acuan pada penelitian.

Penelitian yang dilakukan oleh Stephen Aprius Sutresno (2023) melakukan penelitian tentang dampak penurunan ekonomi global akibat dari resesi. Penelitian ini menggunakan metode *naïve bayes* dan *Support Vector Machine* (SVM). Dataset yang digunakan pada penelitian ini adalah data *crawling* yang bersumber dari media sosial Twitter dari tanggal 1 Januari 2022 s/d 8 Januari 2023. Diperoleh hasil dimana naïve bayes memiliki akurasi 72,5% presisi 83,8% dan *Recall* 79,3% dan SVM dengan hasil berturut turut 79,5%; 78,6%; dan 100%. Dimana SVM memiliki hasil akurasi lebih tinggi daripada naïve bayes (Sutresno,

2023). Pada penelitian ini, akan digunakan topik dan tahun pada dataset sebagai acuan sumber data.

Penelitian oleh Avashlin Moodley dan Dr. Vukosi Marivate (2019) menggunakan *topic modelling* untuk mengetahui topik dan keterkaitannya dari pemilu di Afrika Selatan. Dataset yang digunakan bersumber dari berita *online* yaitu News24. Untuk praprosessing menggunakan *lowercasing, stopword,* streaming, TF-IDF dan menggunakan N-gran. Selanjutnya menggunakan LDA pada kedua pemilu 2014 dan 2019 untuk mengetahui kedua topik tersebut. Hasil dari penelitian tersebut adalah Analisis mengungkap tema terkait korupsi, kebutuhan pertumbuhan ekonomi, dan kelemahan Eskom dalam korpus 2019. Model yang diterapkan pada korpus 2014 mengungkap tema yang berkaitan dengan laporan Nkandla, pertarungan pajak Julius Malema, dan usulan penggabungan DA dan Agang. (Moodley & Marivate, 2019) penelitian ini akan dijadikan acuan dalam penggunaan *preprocessing*.

Tabel 2. 1 Penelitian Terkait

Judul	Penulis	Metode	Kontribusi	Hasil
Deteksi Topik Fashion pada Twitter dengan Latent Dirichlet Allocation	Yupa Umigi Al-khairi, Yudi Wibisono, Budi Laksono Putro	Latent Dirichlet Allocation	Hasil evaluasi dengan *Coherence score*	Hasil dari penelitian ini menunjukkan bahwa PMI terbaik dengan nilai 6.272
Topic modelling pada Sentimen Terhadap Headline Berita *Online* Berbahasa Indonesia Menggunakan LDA dan LSTM	Chairullah Naury, Dhomas Hatta Fudholi, Ahmad Fathan Hidayatullah	Long Short - *term Memory,* dan Latent Dirichlet Allocation	Metode pada penelitan	menghasilkan model dengan nilai akurasi tertinggi sebesar 71.13% dan akurasi terendah sebesar 63.48%.
Topic modelling Penelitian Dosen JPTEI UNY pada Google Scholar	Akhsin Nurlayli, Moch. Ari Nasichuddin	Latent Dirichlet Allocation	Metode pada penelitan	Hasil dari penelitian menunjukkan topik 0 membahas tentang Pendidikan Vokasi, topik 1 membahas tentang

Judul	Penulis	Metode	Kontribusi	Hasil
Menggunakan Latent Dirichlet Allocation				Pengembangan Sistem, topic 2 mebahas tentang interpretasikan berbicara tentang Media Pembelajaran, topik 3 membahas tentang Sistem Pembelajaran di SMK.
Analisis Sentimen Masyarakat Indonesia Terhadap Dampak Penurunan Global Sebagai Akibat Resesi di Twitter	Stephen Aprius Sutresno	*naïve bayes* dan *Support Vector Machine* (SVM)	Topik dan tahun pada dataset	hasil dimana naïve bayes memiliki akurasi 72,5% presisi 83,8% dan *Recall* 79,3% dan SVM dengan hasil berturut turut 79,5%; 78,6%; dan 100%.
Topic modelling of News Articles for Two Consecutive Elections in South Africa	Avashlin Moodley, Dr. Vukosi Marivate	Latent Dirichlet Allocation	Penggunaan Praprocessing dalam penelitian	Analisis mengungkap tema terkait korupsi, kebutuhan pertumbuhan ekonomi, dan kelemahan Eskom dalam korpus 2019. Model yang diterapkan pada korpus 2014 mengungkap tema yang berkaitan dengan laporan Nkandla, pertarungan pajak Julius Malema, dan usulan penggabungan DA dan Agang.

2.2 Sumber Data

Data didapatkan dari dari pengamatan asli peneliti. Pengumpulan informasi menjadi dasar dimana kelangsungan suatu pengumpulan informasi dapat merubah dan mempengaruhi hasil. Untuk hal ini perlu diperhatikan kerapihan dalam pengumpulan data dan kebenarannya. Untuk pengumpulan ini, akan dikumpulkan dari berbagai *website* berita *online* yang memiliki izin pers dalam pembuatan berita. Data yang diambil adalah data *crawling* dengan pencarian “resesi ekonomi” dari tahun 2022 – 2023.

2.3 *Data Mining*

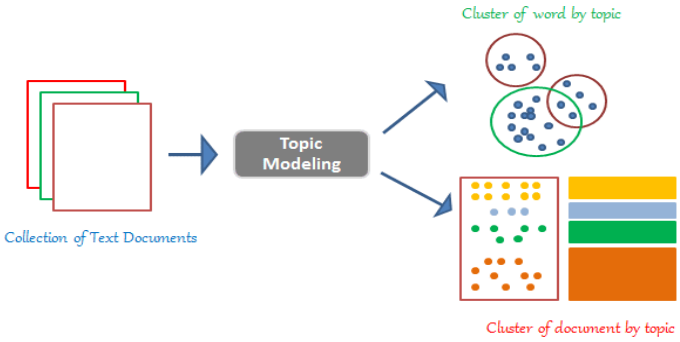
Proses berulang dan interaktif untuk menemukan pola dan model baru yang sempurna, berguna, dan mudah dipahami dalam database yang sangat besar (massive database) merupakan pengertian dari *data mining* (Sikumbang, 2018). *Data mining* digunakan untuk mengetahui trend dari suatu item pada database besar yang dapat digunakan untuk membantu mengambil keputusan. *Data mining* umumnya digunakan untuk prediksi nilai dari sebuah item yang belum dikethui dan memperkirakan nilai yang akan datang. *Data mining* juga digunakan dalam mengklaster data dan objek yang memiliki kemiripan untuk dikelompokkan.

2.4 Resesi Ekonomi

Menurut KBBI, Resesi merupakan keadaan dimana terjadinya keleseuan dalam kegiatan dagang, industry, dan sebagainya menjadi terhenti (KBBI, 2023). Resesi sendiri dikenal dengan kondisi dimana ekonomi di suatu negara mengalami pertumbuhan negatif atau sedang mengalami penurunan berturut – turut dalam satu tahun dalam satu tahun berjalan. Penurunan ekonomi biasanya disertai dengan penurunan harga (deflasi), atau kenaikan harga (inflasi) yang lebih parah dari biasanya. Penyebab resesi lainnya antara lain ketidakseimbangan antara produksi dan konsumsi, penurunan pertumbuhan ekonomi selama dua triwulan berturut-turut, nilai impor lebih besar dari nilai ekspor, dan meningkatnya angka pengangguran (Blandina et al., 2020).

2.5 *Topic Modelling*

Topic modelling adalah salah satu dari teknik *unsupervised machine learning* dimana mampu melakukan analisis sejumlah dokumen, mendeteksi pola kata dan frasa di dalamnya, lalu mengelompokkan kata dan frasa serupa yang paling mewakili sekumpulan dokumen. *Topic modelling* mampu menganalisa *text* dan mengklaster kata dari tiap dokumen (Alfanzar et al., 2020). *Topic modelling* sendiri merupakan teknik *unsupervised* karena tidak memerlukan tag atau label dalam melatihnya. Pada *topic modelling* terdapat dua metode yaitu *Latent Semantic Analysis (LSA)* dan *Latent Dirichlet Allocation (LDA). Topic modelling* digunakan untuk meng-*cluster* atau mengelompokkan kata berdasarkan topik dan mengelompokkan dokumen berdasarkan topik seperti pada gambar 2.1.

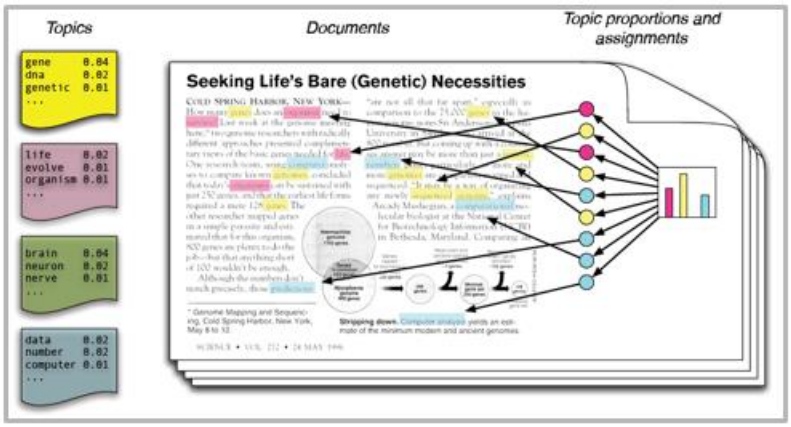


Gambar 2. 1 Proses *Topic*
(sumber: *modelling* https://mandiema.github.io/tdconsulting/)

2.6 *Latent Dirichlet Allocation*

LDA (*Latent Dirichlet Allocation*) adalah metode analisis topik yang dapat digunakan untuk menemukan topik tersembunyi dalam kumpulan dokumen (Setijohatmo et al., 2020). LDA mengasumsikan bahwa setiap dokumen terdiri dari campuran beberapa topik secara proporsional, dan setiap topik terdiri dari campuran

beberapa kata secara proporsional. Ilustrasi pembentukan kata pada topik dapat dilihat pada gambar 2.2.

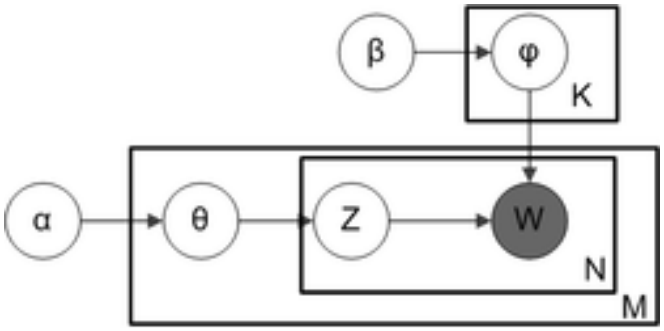


Gambar 2. 2 Ilustrasi Latent Drichlet Allocation
(sumber: https://paperswithcode.com/task/topic-models)

Dalam implementasinya, LDA tidak dapat diterapkan karena variabel *z*, θ dan φ tersembunyi/tidak diketahui. Pada *plate* dapat dilihat untuk gambar yang diarsis adalah variabel yang teramati yaitu *w*, dan variabel lainnya yang tidak diarsis adalah variabel tesembunyi. Dalam proses LDA diatas diasumsikan dengan perulangan generative yang dilakukan untuk setiap dokumen *M* dalam sebuah corpus *D*:

1. Tentukan $\theta_m \sim \text{Dir}(\alpha)$ dimana $m \in \{1, ..., M\}$
2. Tentukan $\varphi_k \sim \text{Dir}(\beta)$ dimana $k \in \{1, ..., K\}$
3. Untuk setiap *N* kata di posisi *m*,*n* dimana $m \in \{1, ..., M\}$, dan $n \in \{1, ..., N_m\}$
    a. Tentukan sebuah topik $Z_{mn} \sim \text{Multinomial}\ (\theta_m)$
    b. Tentukan sebuah kata $W_{mn} \sim$ Multinomial dari $P(W_{nm}|Z_{mn}, \beta)$

Seperti yang ditunjukkan pada gambar berikut, LDA dijelaskan oleh model grafis menggunakan notasi plat pada gambar 2.3.



Gambar 2. 3 Plat Notasi LDA
(sumber: https://en.wikipedia.org/wiki/Latent_Dirichlet_allocation)

Di mana:

- α adalah parameter Dirichlet distribusi dari topik terhadap dokumen
- β adalah parameter Dirichlet distribusi dari kata terhadap topik
- φ_k adalah distribusi kata untuk topik *k*
- θ_m adalah distribusi topik untuk dokumen *m*
- z_{mn} adalah topik untuk kata ke-*n* pada dokumen *m*
- w_{mn} adalah kata pada urutan tertentu ke-*n* pada dokumen *m*
- N adalah himpunan kata yang berada dalam dokumen (dokumen m memiliki N_m kata), dan
- M adalah Kumpulan dari dokumen.

Berdasarkan model, probabilitas dari model pada dokumen adalah sebagai berikut:

$$P(\boldsymbol{W},\boldsymbol{Z},\boldsymbol{\theta},\boldsymbol{\varphi};\alpha,\beta) = \prod_{n=1}^{M} P(\theta_n|\alpha) \prod_{m=1}^{K} P(\varphi_m|\beta) \prod_{o=1}^{N} P(Z_{n,o}|\theta_n) P(W_{n,o}|\varphi_{Zn,o}) \quad (2.1)$$

Pada persamaan (2.1) diasumsikan bahwa topik setiap dokumen mengikuti distribusi *Dirichlet* pada persamaan (2.2):

$$P(\theta|\alpha) = \frac{\Gamma(\alpha)}{\prod_{k=1}^{K}\Gamma(\alpha)} \prod_{k=1}^{K} \Gamma(\theta_k^{\alpha_k - 1}) \quad (2.2)$$

Teruntuk distribusi peluang dari *z* untuk semua dokumen dan topik dalam *terms*, ditunjukkan pada persamaan (2.3) oleh n_m, dan *k* adalah jumlah topik *k* yang termasuk dalam kelompok kata dalam dokumen *m*:

$$P(Z|\theta) = \prod_{m=1}^{M} \prod_{k=1}^{K} \theta_{m.k}^{n_{m,k}} \tag{2.3}$$

Distribusi peluang bersyarat untuk semua *corpus* φ_k juga mengikuti distribusi *Dirichlet* dengan parameter $\varphi_{k,v}$. yang diformulasikan pada persamaan (2.4):

$$P(\varphi|\beta) = \prod_{k=1}^{K} \frac{\Gamma(\beta_{kv})}{\prod_{v=1}^{v} \Gamma(\beta_{kv})} \prod_{v=1}^{V} \varphi_{kv}^{\beta_{kv}-1} \tag{2.4}$$

Dalam menentukan pembagian kata atau distribusi kata dalam pencarian Z_{mn}, LDA menggunakan *gibbs sampling*. *Gibbs sampling* adalah salah satu algoritma keluarga dari *Markov Chain Monte Carlo* (MCMC). *Gibbs sampling* akan men-*sample* satu per satu kata pada setiap dokumen untuk pemberian topik untuk menghubungkan variabel satu dengan variabel lain. Dalam proses *gibbs sampling* akan dilakukan pengelompokan topik dimana proses yang perlu dilakukan yaitu:

1. Distribusi probabilitas topik pada suatu dokumen

Pada proses probabilitas topik pada suatu dokumen, dimisalkan terdapat *n* topik yang muncul pada suatu dokumen *d*, yang ditujukkan pada persamaan (2.5):

$$n_{d,k} + \alpha \tag{2.5}$$

Dimana $n_{d,k}$ adalah jumlah kemunculan topik k yang terjadi pada suatu dokumen, α adalah *hyperparameter* yang mempengaruhi campuran topik.

2. Distribusi probabilitas kata pada suatu topik

Dalam pencarian distribusi probabiltas kata pada topik adalah dengan mengetahui berapa banyak *m* kata *w* muncul pada topik *k* pada corpus. menghasilkan persamaan (2.6):

$$\frac{m_{w,k} + \beta}{\sum_i^V m_{i,k} + W\beta} \tag{2.6}$$

Mirip dengan persamaan sebelumnya, dimana $m_{w,k}$ adalah banyaknya kemunculan kata *w* yang muncul pada topik *k* pada seluruh korpus. *W* adalah banyak term kata yang ada pada dokumen. β adalah *hyperparameter*. Dan $m_{i,k}$ adalah banyaknya kata yang yang tergabung pada topik *k* pada seluruh korpus.

2.7 *Text Processing*

Text processing adalah tahap yang perlu dilakukan sebelum membuat model pelatihan. Proses ini membersihkan data input yang masih kotor, Dimana banyak kata yang tidak perlu digunakan akan dibersihkan. Proses yang akan digunakan pada *text processing* yaitu *Tokenizing, Stopword, Stemming.*

2.7.1 *Tokenizing*

Tokenizing adalah proses membagi teks menjadi unit yang lebih kecil, yang berupa kata atau frasa. *Tokenizing* adalah langkah penting dalam *text mining* karena memungkinkan untuk memisahkan setiap kata ke dalam unit-unit kecil dalam suatu array atau term (Togatorop et al., 2021). Pada *tokenizing* juga diawali dengan *lowercasing* dimana teks yang ada akan diubah menjadi huruf kecil untuk memberikan hasil lebih maksimal.

2.7.2 *Stopword*

Stopword adalah kata yang muncul dalam secara berulang pada teks tetapi tidak memberikan informasi penting atau spesifik. Seperti kata depan (di, pada, dari), kata hubung (dan, atau), kata ganti (saya, dia, mereka), dll. Dalam *text mining*, kata-kata berhenti biasanya dihapus dari teks sebelum analisis lebih lanjut. Tujuannya adalah untuk mengurangi ukuran data, mempercepat proses pengindeksan, dan meningkatkan relevansi hasil pencarian (Nugraha et al., 2017).

2.7.3 *Stemming*

Stemming adalah proses menghasilkan perubahan morfologis dari akar kata. *Stemming* adalah teknik yang digunakan untuk menormalkannya agar lebih mudah untuk diproses.

2.8 Membangun *Corpus*

Membangun *corpus* adalah tahap yang dilakukan setelah *text processing*. Tahap ini akan membangun dan memperkaya *corpus* dan *dictionary* yang nantinya digunakan pada model. Dua tahap di dalam pembangunan *corpus* adalah *N-gram* dan TF-IDF

2.8.1 *N-gram*

N-gram adalah salah satu teknik dalam *text mining* yang digunakan untuk mengukur kemiripan antara dua teks atau dokumen. *N-gram* adalah kumpulan n buah kata yang berurutan dalam sebuah teks. Misalnya, *n-gram* dengan n=2 disebut *bigram*, *n-gram* dengan n=3 disebut *trigram*, dan seterusnya. Untuk menghitung kemiripan antara dua teks menggunakan *n-gram*, kita dapat menghitung jumlah *n-*

gram yang sama di kedua teks dan membaginya dengan jumlah total *n-gram* di kedua teks(Anugrah, 2021). Semakin tinggi nilai kemiripan, semakin mirip kedua teks tersebut.

2.8.2 TF-IDF

Nama lengkap TF-IDF adalah *Term Frequency Inverse Document Frequency*, yaitu metode penghitungan bobot kata dalam sebuah dokumen. TF-IDF dapat digunakan untuk menentukan tingkat kepentingan suatu kata dalam suatu dokumen atau korpus. Rumus yang digunakan untuk menghitung TF-IDF menggunakan persamaan (2.7):

$$\mathrm{tf-idf(t,d)} = \mathrm{tf(t,d)} \times \mathrm{idf(t)} \tag{2.7}$$

Dimana t adalah kata, d adalah dokumen, tf(t,d) adalah frekuensi kata t dalam dokumen d, tf(t,d) dihitung dengan rumus (2.8) :

$$\mathrm{tf(t,d)} = \frac{\mathrm{f_{t,d}}}{\Sigma \mathrm{f_{t',d}}} \tag{2.8}$$

idf(t) adalah frekuensi inversi kata t dalam dokumen, dihitung dengan rumus (2.9):

$$\mathrm{idf}(t) = \log \frac{n}{\mathrm{df}(t)} \tag{2.9}$$

dimana n adalah jumlah total dokumen dalam korpus dan df(t) adalah jumlah dokumen dalam korpus yang mengandung kata t.

2.9 *Coherence Score*

Kita dapat menggunakan *coherence score* dalam *topic modelling* untuk mengukur seberapa mudah suatu topik dapat dipahami oleh manusia. Dalam hal ini, *coherence score* digunakan sebagai validasi dari hasil *topic modelling*. Singkatnya, skor konsistensi mengukur seberapa mirip kata-kata ini satu sama lain. Salah satu dari model evaluasi yang digunakan adalah *Umass* dan *Cv coherence score*. *Umass* menghitung seberapa sering dua kata muncul yaitu w_i dan w_j bersamaan pada sebuah corpus. *Umass* didefinisikan pada persamaan (2.10):

$$Umass(w_i, w_j) = \log \frac{P(w_i, w_j) + 1}{P(w_i)} \tag{2.10}$$

Dimana $P(w_i, w_j)$ adalah banyaknya kemunculan kata w_i dan w_j muncul pada sebuah dokumen. Dan $P(w_i)$ adalah probabibilitas kata w_i w_j, muncul pada sebuah model latih

Model evaluasi *Cv coherence score* menghitung seberapa sering dua kata muncul yaitu w_i dan w_j bersamaan pada sebuah korpus. *Cv* didefinisikan pada persamaan (2.10):

$$Cv(w_i, w_j) = \log \frac{P(w_i, w_j) + 1}{P(w_i)P(w_j)} \tag{2.11}$$

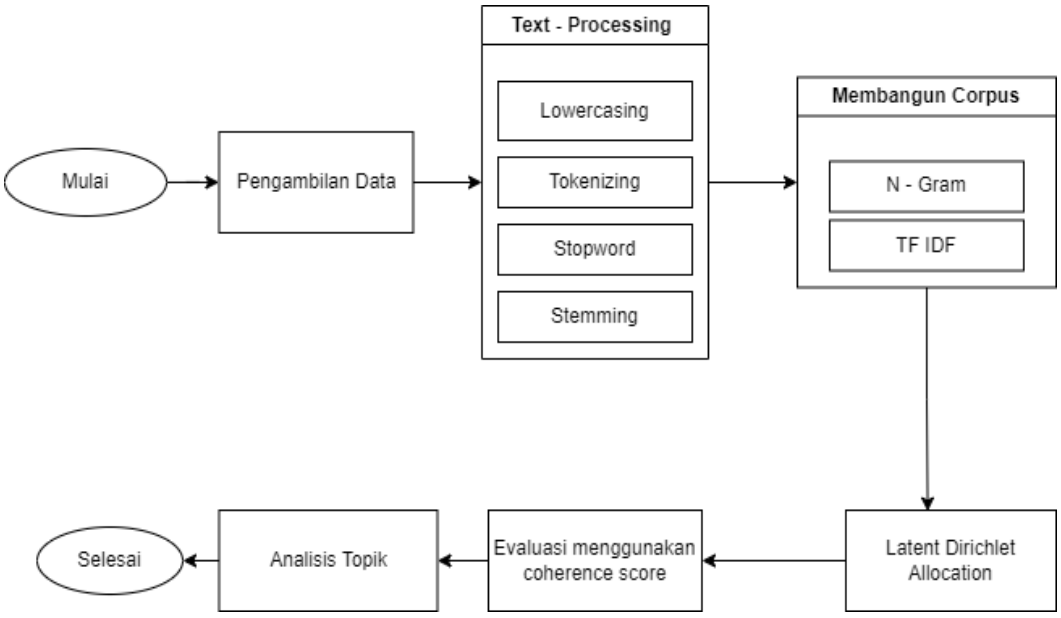
Dimana $P(w_i, w_j)$ adalah banyaknya kemunculan kata w_i dan w_j muncul pada sebuah dokumen. Dan $P(w_i), P(w_j)$ adalah probabibilitas kata w_i w_j, muncul pada sebuah model latih (Trenquier, 2018).

BAB III

DESAIN PENELITIAN

3.1 Alur Penelitian

Data yang sudah diolah akan di proses seperti desain alogaritma yang akan digunakan pada diagram gambar 3.1:



Gambar 3. 1 Diagram alur proses penelitian

3.2 Pengambilan data

Dalam penelitian ini, peneliti mengumpulkan data dari *web crawling* dan *scrapping* secara *online* melalui situs berita seperti detik.com. Data yang dikumpulkan akan di simpan dalam bentuk csv untuk memudahkan pemrosesan pada program python, dengan batasan datanya adalah data yang dikumpukan menggunakan *topic of interest* mengenai resesi dan hal – hal terkait mengenai resesi dan untuk waktunya adalah mulai dari tahun 2022 hingga 2023. Dalam dataset yang

digunakan akan terdiri dari judul, kategori, waktu terbit, dan sumber detik ditampilkan pada tabel 3.1:

Tabel 3. 1 *Dataset*

Judul	Kategori	Waktu terbit	Sumber
Harga Terus Fluktuatif, Emas Masih Tepat untuk Investasi?	detikFinance	27 Mar 2023 09:37 WIB	https://finance.detik.com/berita-ekonomi-bisnis/d-6639619/harga-terus-fluktuatif-emas-masih-tepat-untuk-investasi
Populasi Jepang 'Anjlok', 10 Negara Ini Juga Catat Angka Kelahiran Terendah Dunia	detikHealth	26 Mar 2023 19:01 WIB	https://health.detik.com/berita-detikhealth/d-6639123/populasi-jepang-anjlok-10-negara-ini-juga-catat-angka-kelahiran-terendah-dunia
'Biang Kerok' Populasi Negeri Sakura Anjlok, Pasangan Jepang Kurang Romantis?	detikHealth	26 Mar 2023 16:30 WIB	https://health.detik.com/berita-detikhealth/d-6638936/biang-kerok-populasi-negeri-sakura-anjlok-pasangan-jepang-kurang-romantis
Cek! Ini Daftar Krisis Keuangan Terbesar dalam 40 Tahun Terakhir	detikFinance	26 Mar 2023 10:57 WIB	https://finance.detik.com/berita-ekonomi-bisnis/d-6638447/cek-ini-daftar-krisis-keuangan-terbesar-dalam-40-tahun-terakhir
Dua Bank Besar Dunia Kolaps, RI Bisa Kena Getahnya?	detikFinance	23 Mar 2023 22:32 WIB	https://finance.detik.com/moneter/d-6634775/dua-bank-besar-dunia-kolaps-ri-bisa-kena-getahnya
Warga Ogah Punya Anak, Korea Selatan Catat Rekor Jumlah Perkawinan Teranjlok	detikHealth	21 Mar 2023 07:30 WIB	https://health.detik.com/berita-detikhealth/d-6629840/warga-ogah-punya-anak-korea-selatan-catat-rekor-jumlah-perkawinan-teranjlok
Ekonom Sumut Minta Pemerintah Tak Buru-buru Setop Impor Pakaian Bekas	detikSumut	20 Mar 2023 23:08 WIB	https://www.detik.com/sumut/bisnis/d-6628519/ekonom-sumut-minta-pemerintah-tak-buru-buru-setop-impor-pakaian-bekas
Industri Bisnis Pernikahan Diprediksi Makin Cerah Tahun Ini	detikSumut	19 Mar 2023 03:02 WIB	https://www.detik.com/sumut/bisnis/d-6626664/industri-bisnis-pernikahan-diprediksi-makin-cerah-tahun-ini
Pakai Aplikasi Ini Bisa Beli Saham di Wall Street Mulai Rp 5.000	detikFinance	16 Mar 2023 22:42 WIB	https://finance.detik.com/bursa-dan-valas/d-6623519/pakai-aplikasi-

Judul	Kategori	Waktu terbit	Sumber
			ini-bisa-beli-saham-di-wall-street-mulai-rp-5000
Tren Penjualan Laptop Asus: Consumer Turun dan Gaming Naik	detikInet	16 Mar 2023 20:51 WIB	https://inet.detik.com/consumer/d-6623204/tren-penjualan-laptop-asus-consumer-turun-dan-gaming-naik
LPS Pastikan Bank Silicon Valley Bangkrut Nggak Ngaruh ke RI	detikFinance	16 Mar 2023 18:45 WIB	https://finance.detik.com/moneter/d-6623061/lps-pastikan-bank-silicon-valley-bangkrut-nggak-ngaruh-ke-ri
Bahaya! Bangkrutnya Bank Silicon Valley Bisa Picu Resesi Global	detikFinance	16 Mar 2023 15:05 WIB	https://finance.detik.com/moneter/d-6622452/bahaya-bangkrutnya-bank-silicon-valley-bisa-picu-resesi-global

Tabel 3.1 menampilkan hasil dari proses pengambilan data yang telah dilakukan, Dimana dalam isi sebuah artikel, diperoleh 4 kolom yaitu judul, kategori, dan waktu terbit, data pada tabel yang telah diperoleh akan melakukan *crawling* menuju link tertera pada sumber untuk mengambil isi dari artikel tersebut.

3.3 Text Processing

Pada *text processing*, data yang telah dikumpulkan akan diproses untuk diolah sebelum dapat digunakan metode LDA. Tahap ini membersihkan dan menstandarisasi teks dari berbagai sumber dokumen. *Text processing* yang digunakan pada penelitian ini menggunakan *tokenizing, stopword, stemming*, TF-IDF dan *n-gram*. Pada proses *tokenizing*, kalimat akan diubah menjadi huruf kecil dengan *lowercasing*.

Tabel 3. 2 Proses *Tokenizing*

Sebelum *Tokenizing*	**Sesudah *Tokenizing***
Harga emas setiap harinya cenderung bergerak fluktuatif. Hal ini terjadi karena masih adanya bayang-bayang resesi, perang Rusia-Ukraina sampai krisis perbankan	['harga', 'emas', 'setiap', 'harinya', 'cenderung', 'bergerak', 'fluktuatif.', 'hal', 'ini', 'terjadi', 'karena', 'masih', 'adanya', 'bayang-bayang', 'resesi,', 'perang', 'rusia-ukraina', 'sampai', 'krisis', 'perbankan']

Hasil proses *tokenizing* ditampilkan pada tabel 3.2 dimana teks yang awalnya merupakan satu paragraf string akan di pisah dengan fungsi pada python menjadi kumpulan kata. Proses yang bersamaan dilakukan fungsi *lowercasing.* Fungsi ini digunakan untuk menghilangkan huruf kapital pada kata.

Teks yang telah di lakukan *tokenizing* selanjutnya diproses untuk dihilangkan kata hubung yang digunakan dengan *stopword.* Kata-kata yang dirpsoes di *stopword* akan menghilangkan teks yang berulang dan tidak penting untuk memaksimalkan pencarian nantinya.

Tabel 3. 3 Proses *Stopword*

Sebelum *Stopword*	**Sesudah *Stopword***
['harga', 'emas', 'setiap', 'harinya', 'cenderung', 'bergerak', 'fluktuatif.', 'hal', 'ini', 'terjadi', 'karena', 'masih', 'adanya', 'bayang-bayang', 'resesi,', 'perang', 'rusia-ukraina', 'sampai', 'krisis', 'perbankan']	['harga', 'emas', '', 'harinya', 'cenderung', 'bergerak', 'fluktuatif.', '', '', 'terjadi', '', '', 'adanya', 'bayang-bayang', 'resesi,', 'perang', 'rusia-ukraina', '', 'krisis', 'perbankan']

Stopword dapat menggunakan *library* NLTK (*natural language toolkit*) untuk menghilangkan kata yang tidak penting dalam teks. Dalam penggunaan NLTK, terdapat kekurangan dimana masih belum bagusnya dalam mengeleminasi kata dalam Bahasa Indonesia. Hal ini dapat diatasi dengan menggunakan *library* untuk Bahasa Indonesia yaitu *library* sastrawi. Ditampilkan pada tabel 3.3 dimana proses *stopword* menghilangkan kata-kata yang umum digunakan seperti kata

setiap, hal, ini, karena, dan sampai. Kata tersebut akan dihapus pada Kumpulan kata.

Selanjutnya dilakukan *stemming* pada *preprocessing*. Proses ini akan mengubah kata menjadi kata dasar atau akar dari kata tersebut.

Tabel 3. 4 Proses *Stemming*

Sebelum *Stemming*	**Sesudah *stemming***
['harga', 'emas', '', 'harinya', 'cenderung', 'bergerak', 'fluktuatif.', '', '', 'terjadi', '', '', 'adanya', 'bayang-bayang', 'resesi,', 'perang', 'rusia-ukraina', '', 'krisis', 'perbankan']	['harga', 'emas', '', 'hari', 'cenderung', 'gerak', 'fluktuatif', '', '', 'jadi', '', '', 'ada', 'bayang', 'resesi', 'perang', 'rusia-ukraina', '', 'krisis', 'perban']

Pada tabel 3.4 beberapa kata yang sebelumnya memiliki kata imbuhan seprti kata harinya akan diubah menjadi kata aslinya yaitu hari, lalu pada kata bergerak menjadi kata gerak dan kata bayang-bayang yang merupakan kata berulang diproses menjadi kata bayang.

3.4 Membangun Corpus

N-gram yang digunakan nantinya adalah *bigram* yang membuat satu buah kata bisa terdiri dari dua karakter agar tidak merubah makna dari kata tersebut. Selanjutnya dilakukan tf-idf untuk menghitung frekuensi dari kata di sebuah dokumen. Proses pertama yang perlu dilakukan tf – idf adalah menyediakan dokumen terlebih dahulu.

D1: Dengan kondisi ekonomi global yang diperkirakan masih dihantui resesi, maka emas menjadi instrumen investasi *safe haven*.

D2: Harga emas setiap harinya cenderung bergerak fluktuatif. Hal ini terjadi karena masih adanya bayang-bayang resesi, perang Rusia-Ukraina sampai krisis perbankan

D3: Menteri Keuangan Sri Mulyani mengungkap, para menteri keuangan dari negara anggota G20 bersama gubernur bank sentral telah membahas ancaman resesi 2023.

Langkah pertama dalam TF - IDF adalah data diproses untuk cleaning,*stopword*, dan stemming yang selanjutnya di *tokenizing*.

Document 1: [kondisi, ekonomi, global, diperkirakan, dihantui, resesi, emas, instrumen, investasi, safe, haven]

Document 2: [Harga, emas, harinya, cenderung, bergerak, fluktuatif, Hal, terjadi, adanya, bayang-bayang, resesi, perang, Rusia-Ukraina, krisis, perbankan]

Document 3: [Menteri, Keuangan, Sri, Mulyani, mengungkap, menteri, keuangan, negara, anggota, G20, bersama, gubernur, bank, sentral, membahas, ancaman, resesi, 2023]

Langkah kedua, hitung jumlah *Term Frequency* (TF).

TF mewakili frekuensi *term* dalam dokumen. Proses dilakukan dengan menghitung berapa kali setiap *term* muncul di setiap dokumen. Perhitungan ditampilkan pada tabel 3.5:

Tabel 3. 5 Term Frequency (TF)

Term	D0	D1	D2
kondisi	1		
ekonomi	1		
global	1		
diperkirakan	1		
dihantui	1		
resesi	1	1	1
emas	1	1	
instrumen	1		
investasi	1		
save	1		
haven	1		
harga		1	
harinya		1	
cenderung		1	
bergerak		1	
fluktuatif		1	
hal		1	
terjadi		1	
adanya		1	
bayang-bayang		1	
perang		1	
rusia-ukraina		1	
krisis		1	
perbankan		1	
menteri			2
keuangan			2
sri			1
mulyani			1
mengungkap			1
menteri			1
keuangan			1
negara			1
anggota			1
g20			1
bersama			1
gubernur			1
bank			1
sentral			1
membahas			1
ancaman			1
2023			1

Proses penentuan *term frequency* ditampilkan pada tabel 3.5 dimana terdapat tiga dokumen yang digunakan. Ketiga dokumen D0, D1, D2 memiliki kata resesi didalamnya, selanjutnya pada dokumen D0 dan D1 memilki kata emas didalamnya dan pada dokumen D2, terdapat kata menteri dan keuangan yang

muncul sebanyak dua kali. Selanjutnya proses normalisasi tf(t,d) dengan menggunakan rumus (2.2),

$$\text{tf("kondisi", D0)} = \frac{banyak\ term\ kondisi\ muncul\ pada\ dokumen}{total\ term\ yang\ muncul\ pada\ dokumen} \quad (3.1)$$

sehingga diperoleh hasil tf dengan nilai 0,091. Hasil normalisasi *term frequency* selanjutnya ditampilkan pada tabel 3.6.

Tabel 3. 6 Normalisasi *Term Frequency* (TF)

Term	**D0**	**D1**	**D2**
kondisi	0,091	0	0
ekonomi	0,091	0	0
global	0,091	0	0
diperkirakan	0,091	0	0
dihantui	0,091	0	0
resesi	0,091	0,067	0,056
emas	0,091	0,067	0
instrumen	0,091	0	0
investasi	0,091	0	0
save	0,091	0	0
haven	0,091	0	0
Harga	0	0,067	0
harinya	0	0,067	0
cenderung	0	0,067	0
bergerak	0	0,067	0
fluktuatif	0	0,067	0
Hal	0	0,067	0
terjadi	0	0,067	0
adanya	0	0,067	0
bayang-bayang	0	0,067	0
perang	0	0,067	0
Rusia-Ukraina	0	0,067	0
krisis	0	0,067	0
perbankan	0	0,067	0
Menteri	0	0	0,111
Keuangan	0	0	0,111
Sri	0	0	0,056
Mulyani	0	0	0,056
mengungkap	0	0	0,056
negara	0	0	0,056
anggota	0	0	0,056
G20	0	0	0,056
bersama	0	0	0,056
gubernur	0	0	0,056
bank	0	0	0,056
sentral	0	0	0,056
membahas	0	0	0,056
ancaman	0	0	0,056
2023	0	0	0,056

Tabel 3.6 menampilkan hasil dari normalisasi TF yang telah dilakukan pada tiga dokumen sebelumnya. Terdapat beberapa hasil yang berbeda seperti pada kata menteri dan keuangan memiliki nilai TF yaitu 0,111 karena jumlah kemunculannya pada dokumen lebih dari satu.

Selanjutnya dilakukan perhitungan untuk mencari IDF. Menggunakan perhitungan rumus (2.3), perhitungan idf(t), perlu mencari frekuensi kemunculan term pada semua dokumen atau korpus df(t). Seperti pada kata resesi dimana memiliki df(t) berjumlah 3.

$$\text{Idf}(\text{“}kondisi\text{”}) = \log \frac{jumlah\ dokumen}{jumlah\ term\ kondisi\ di\ semua\ dokumen} = \log\left(\frac{3}{1}\right) = 0{,}477 \quad (3.2)$$

Selanjutnya dilakukan pencarian tf-idf dengan formula (2.7), didapatkan hasil:

$$tf - idf(\text{“}kondisi, d0) = 0{,}091 * 0{,}477 = 0.04341803418 \quad (3.3)$$

Hasil tersebut akan digunakan sebagai bobot kata kondisi pada dokumen D0. Proses TF=IDF ditampilkan pada tabel 3.7.

Tabel 3. 7 TF-IDF

TERM	DF	IDF	TF-IDF		
			D0	D1	D2
kondisi	1	0,477	0,043	0	0
ekonomi	1	0,477	0,043	0	0
global	1	0,477	0,043	0	0
diperkirakan	1	0,477	0,043	0	0
dihantui	1	0,477	0,043	0	0
resesi	3	0	0	0	0
emas	2	0,176	0,016	0,012	0
instrumen	1	0,477	0,043	0	0
investasi	1	0,477	0,043	0	0
save	1	0,477	0,043	0	0
haven	1	0,477	0,043	0	0
Harga	1	0,477	0	0,032	0
harinya	1	0,477	0	0,032	0
cenderung	1	0,477	0	0,032	0
bergerak	1	0,477	0	0,032	0

TERM	DF	IDF	TF-IDF		
			D0	D1	D2
fluktuatif	1	0,477	0	0,032	0
Hal	1	0,477	0	0,032	0
terjadi	1	0,477	0	0,032	0
adanya	1	0,477	0	0,032	0
bayang-bayang	1	0,477	0	0,032	0
perang	1	0,477	0	0,032	0
Rusia-Ukraina	1	0,477	0	0,032	0
krisis	1	0,477	0	0,032	0
perbankan	1	0,477	0	0,032	0
Menteri	1	0,477	0	0	0,053
Keuangan	1	0,477	0	0	0,053
Sri	1	0,477	0	0	0,027
Mulyani	1	0,477	0	0	0,027
mengungkap	1	0,477	0	0	0,027
negara	1	0,477	0	0	0,027
anggota	1	0,477	0	0	0,027
G20	1	0,477	0	0	0,027
bersama	1	0,477	0	0	0,027
gubernur	1	0,477	0	0	0,027
bank	1	0,477	0	0	0,027
sentral	1	0,477	0	0	0,027
membahas	1	0,477	0	0	0,027
ancaman	1	0,477	0	0	0,027
2023	1	0,477	0	0	0,027

Tabel 3.7 memberikan hasil dari perkalian antara tf-idf dimana proses menghasilkan nilai bobot pada setiap dokumen. Perbedaan nilai bobot pada dokumen yang sama diakibatkan karena kemunculan pada dokumen terjadi lebih dari satu kali yaitu pada kata menteri dan keuangan dimana diperoleh 0,053. Pada D0 diperoleh hasil 0,043 dan 0,032 pada dokumen D1 pada setiap kata selain kata resesi. Kata resesi memiliki bobot bernilai 0 karena kata tersebut ada pada setiap dokumen sehingga mengurangi nilai bobotnya. Hal ini juga dilihat pada kata emas yang memiliki bobot paling kecil yaitu 0,016 pada dokumen D0 dan 0,012 pada D1.

3.5 Latent Dirichlet Allocation (LDA)

Setelah data telah melewati tahap text processing, tahap selanjutnya adalah pembuatan model LDA. Proses LDA pada gensim dapat diproses seperti pada rumus untuk mendapatkan persamaan (2.1) dengan penentuan parameter dirichlet yaitu *alpha* dan *beta* diikuti dengan penetuan besar topik atau K. Pada sampel selanjutnya akan digunakan parameter dengan nilai $\alpha = 0.1$, $\beta = 0.1$, $K = 3$.

Kumpulan dokumen yang telah di tokenisasi akan diubah dalam bentuk korpus yang selanjutnya diinisialisasi topik untuk semua kata dalam dokumen secara random (topik 1 adalah K1, dst). Berikut adalah simulasi untuk 3 dokumen (D1, D2, dan D3) dan topik tersembunyi.

Tabel 3. 8 Inisialisasi Topik Secara *Random*

No	D1	K	D2	K	D3	K
1	resesi	1	resesi	3	resesi	1
2	bank	1	resesi	2	resesi	2
3	bank	2	bank	1	bank	1
4	bank	1	bank	2	bank	1
5	investasi	3	bank	3	bank	2
6	ekonomi	2	bank	3	bank	1
7	minyak	1	bank	3	bank	1
8	minyak	2	investasi	1	bank	1
9	minyak	3	ekonomi	1	bank	2
10	minyak	2	ekonomi	2	bank	3
11	minyak	2	ekonomi	1	investasi	2
12	rusia	1	ekonomi	3	krisis	2
13	rusia	3	minyak	2	ekonomi	1

Tabel 3.8 dilakukan pemberian topik secara random pada teks yang telah diproses pada setiap dokumen. Karena pada sebelumnya telah ditentukan topik sebanyak 3, maka pada setiap kata akan menjadi kata pada topik satu, dua dan tiga. Dari tabel 3.8 dapat diketahui distribusi topik terhadap kata dan distribusi dokumen terhadap topik yang ditampilkan pada tabel 3.9 dan 3.10:

Tabel 3. 9 Distribusi Topik Pada Kata

	K1	**K2**	**K3**
bank	8	4	4
ekonomi	3	2	1
investasi	1	1	1
krisis	0	1	0
minyak	1	4	1
resesi	2	2	1
usia	1	0	1

Persebaran kata terhadap topik ditampilkan pada tabel 3.9 dimana kata pada topik K1 memiliki kata bank cukup banyak dibandingkan pada K2 dan K3, kata ekonomi memiliki penempatan pada topik K1 sebanyak tiga kali, lalu dilanjutkan pada K2 sebanyak dua kali dan K3 sebanyak sekali. Pada kata investasi, mendapat penempatan yang sama pada setiap topik. Kata krisis yang muncul hanya satu kali pada dokumen D2 saja mendapat asumsi sebagai topik K2. Kata minyak memiliki penempatan pada K2 terbesar sebanyak 4 kali disbandingkan pada K1 dan K3. Kata resesi memiliki penempatan sama pada K1 dan K2 sebanyak dua kali dan hanya sekali pada K3. Kata usia memiliki penempatan sekali pada K1 dan K3 dan bernilai 0 karena hanya dua kali kata usia muncul pada semua dokumen.

Tabel 3. 10 Distribusi Dokumen Terhadap Topik

	D1	**D2**	**D3**
K1	5	4	7
K2	5	4	5
K3	3	5	1

Tabel 3.10 menampilkan distribusi dokumen terhadap topik yang telah ditetapkan secara random sebelumnya. Terlihat pada dokumen D1 memiliki persebaran topik K1 dan K2 sebanyak 5 dan K3 sebanyak 3. Pada dokumen D2 memiliki persebaran topik K1 dan K2 sebanyak 4 dan K3 sebanyak 5 kata. Pada

dokumen D3 memiliki persebaran topik K1 paling banyak dengan jumlah 7, K2 sebanyak 5 dan K3 sebanyak 1.

Selanjutnya menggunakan persamaan (2.5) dan (2.6) untuk mendapat probablitas kata terhadap topik:

$$P(Z = K|z, w, d) = n_{d,k} + \alpha \times \frac{m_{w,k} + \beta}{\sum_i^V m_{i,k} + W\beta} \quad (3.4)$$

Pada simulasi ini, akan di lakukan sample kata pertama pada dokmen 1 yaitu “resesi”.

$$P(Z = K_1|z, resesi, d_1) = 5 + 0{,}1 \times \frac{2 + 0{,}1}{16 + 7 * 0{,}1} = 0.641 \quad (3.5)$$

$$P(Z = K_2|z, resesi, d_1) = 5 + 0{,}1 \times \frac{2 + 0{,}1}{14 + 7 * 0{,}1} = 0.728 \quad (3.6)$$

$$P(Z = K_3|z, resesi, d_1) = 3 + 0{,}1 \times \frac{1 + 0{,}1}{9 + 7 * 0{,}1} = 0.351 \quad (3.7)$$

Dari hasil diatas dapat dilihat bahwa topik 2 memiliki hasil yang lebih besar, sehingga kata “resesi” pada dokumen 1 merupakan topik 2.

Tabel 3. 11 Kata Z Setelah *Gibbs Sampling* Pada D1

No	D1	K	D2	K	D3	K
1	resesi	2	resesi	3	resesi	1
2	bank	1	resesi	2	resesi	2
3	bank	2	bank	1	bank	1
4	bank	1	bank	2	bank	1
5	investasi	3	bank	3	bank	2
6	ekonomi	2	bank	3	bank	1
7	minyak	1	bank	3	bank	1
8	minyak	2	investasi	1	bank	1
9	minyak	3	ekonomi	1	bank	2
10	minyak	2	ekonomi	2	bank	3
11	minyak	2	ekonomi	1	investasi	2

No	D1	K	D2	K	D3	K
12	rusia	1	ekonomi	3	krisis	2
13	rusia	3	minyak	2	ekonomi	1

Pada tabel 3.11, pada kata resesi di dokumen D1 mengalami perubahan pada penempatan topik yang sebelumnya dilakukan secara random yaitu sebagai topik 1, kini setelah diproses menggunakan *gibbs sampling* menjadi topik 2. Proses distribusi topik terhadap kata dan dokumen terhadap topik dilakukan lagi menyesuiakan data yang telah diperbarui.

Tabel 3. 12 Hasil Penyesuaian Distribusi Topik Terhadap Kata

	K1	K2	K3
bank	8	4	4
ekonomi	3	2	1
investasi	1	1	1
krisis	0	1	0
minyak	1	4	1
resesi	1	3	1
usia	1	0	1

Tabel 3.12 merupakan penyesuaian dari distribusi yang dilakukan sebelumnya dalam topik terhadap kata. Kata resesi pada dokumen D0 yang telah ditetapkan menjadi topik 2, sehingga kata resesi pada topik K1 berkurang dan pada topik K2 bertambah. Selanjutnya penyesesuaian distribusi dokumen terhadap topik ditampilkan pada tabel 3.13:

Tabel 3. 13 Hasil Penyesuaian Distribusi Dokumen Terhadap Topik

	D1	D2	D3
K1	4	4	7
K2	6	4	5
K3	3	5	1

Tabel 3.13 menyesuiakan perubahan yang telah terjadi pada kata resesi, kini pada dokumen D1, mengalami perubahan Dimana kata resesi sebelumnya

merupakan topik K1 kini menjadi K2, sehingga K1 pada dokumen berkurang satu dan K2 pada dokumen D1 bertambah satu. Proses pada diatas akan dilakukan terus menerus hingga semua semua kata pada corpus sudah dilakukan proses *gibbs sampling* dan mendapatkan topik yang sesuai berdsarkan probabilitas topik LDA.

Metode LDA pada penelitian ini akan menggunakan *library* python yaitu gensim. Pemanggilan LDA dengan *library* gensim berisi parameter berupa *id2word* yang berasal dari Kumpulan kata yang di map dengan ID, parameter corpus yang merupakan kumpulan ID dengan jumlah yang muncul pada setiap dokumen, parmater *num_topic* untuk menentukan berapa banyak topik yang akan diekstrak dalam model pelatihan, Output yang dilkeuarkan dari LDA berupa Kumpulan topik yang akan ditampilkan dengan grafik dan *word cloud*.

3.6 Evaluasi

Model LDA akan di evaluasi menggunakan metode pemrosesan yaitu *coherence score*, *Umass coherence* dan *Cv Coherence score* dimuat sebagai parameter dalam menentukan *coherence score*. Analisis dilakukan dengan melihat performa dari *Umass* dan *Cv*. Dimana batas *coherence score* untuk *Umass* adalah -14 sampai 14, Menurut pengukuran *Umass coherence score*, koherensi topik secara global menurun saat K meningkat, dan model dikatakan baik jika *score* mendekati 0 berarti koherensi semakin sempurna. Sedangkan pada *Cv coherence score* akan mengalami penaikan saat K meningkat, dan model dikatakan baik jika skor *Cv* yang positif dan tinggi menandakan kualitas koherensi topik yang optimal. (Trenquier, 2018).

3.7 Analisis Topik

Pada analisis topik akan ditampilkan persebaran kata dalam topik dengan visualisasi *wordcloud,* dengan mengetahui kata paling besar probabilitasnya dapat diketahui fenomena atau berita menjadi isu utama dalam resesi. Selanjutnya dilakukan interpretasi untuk mengetahui label atau hal yang dibahas pada topik tersebut. Dalam mencari hubungan penyebab resesi dengan topik dapat menggunakan parameter resesi yaitu ekonomi, negatif, fluktuasi, inflasi, produk domestik bruto, produksi, distribusi, konsumsi, investasi, phk, krisis, turun, deflasi. Parameter selanjutnya akan di proses pada setiap topik untuk dicari muncul atau tidak lalu diranking. Dalam ranking akan menggunakan rumus

$$rank = \frac{n(kata)}{n(resesi)} \times 100\% \tag{3.8}$$

dimana n(kata) adalah banyak kata parameter yang muncul pada topik, n(populasi) adalah total parameter resesi. Dari proses diatas diperoleh urutan topik dengan pembahasan resesi dari terbanyak ke sedikit.

BAB IV

HASIL DAN PEMBAHASAN

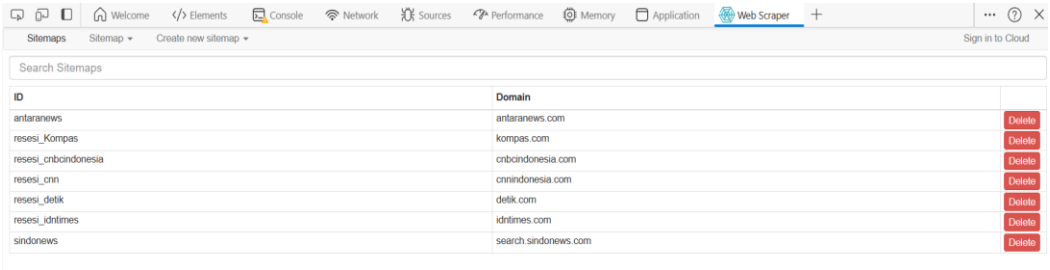
4.1 Pengambilan Data

Pada pengambilan data ini bersumber dari media berita *online* yang terdaftar pada dewan pers dan telah memiliki sertifikat. Dewan pers adalah Lembaga yang melindungi kemerdekaan pers dari campur tangan pihak lain dan Melakukan pengkajian untuk pengembangan kehidupan pers. Setifikasi Perusahaan pers yang dikeluarkan oleh dewan pers dijadikan acuan dikarenakan telah terverifikasi secara faktual dan membuktikan bahwa perusahaan telah mematuhi peraturan dan undang-undang turunan yang telah disetujui oleh komunitas pers. Berdasar pada hal tersebut, media yang dijadikan sebagai dataset yaitu CNN Indonesia, Detik.com, Kompas.com, Antaranews, Idntimes, CNN Indonesia adalah salah satu perusahaan pers yang dimiliki oleh Trans Media yang bekerja sama dengan Warner Bros. Detik.com juga merupakan salah satu dari media *online* yang dimiliki oleh Trans media dengan anak perusahaan CT Corp sebagai pemilik media digital. Kompas.com merupakan salah satu dari media digital yang dimiliki perusahan PT Kompas Media Nusantara. Antaranews adalah agensi berita yang dimiliki oleh pemerintah Indonesia. Idntimes adalah salah satu media digital yang dimiliki oleh IDN media.

Tabel 4. 1 Sumber Berita Data

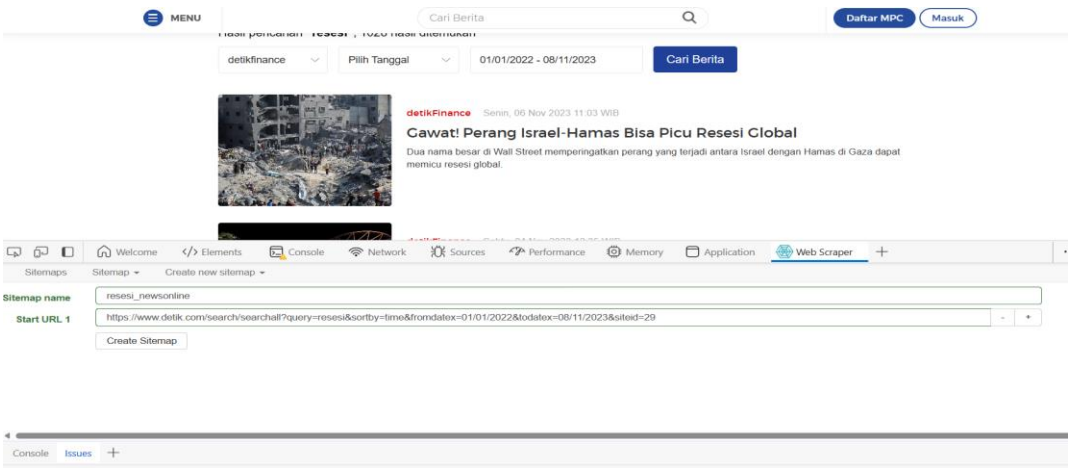
Sumber Berita	**Jumlah**
CNN Indonesia	2331
Detik.com	1028
Kompas.com	313
Antaranews	1500
Idntimes	304

Pada tabel 4.1, ditampilkan jumlah dari dataset *crawling* pada berita *online* yang digunakan. *Crawling* menggunakan *add-ons* atau ekstensi dari Microsoft Edge yaitu *web scrapper*. *Web scraper* dapat digunakan setelah melakukan inspect pada situs yang akan diproses seperti pada gambar 4.1. Penambahan *sitemap* dengan membuat *sitemap* baru seperti pada gambar 4.2.



Gambar 4. 1 Tampilan Muka Pada *Web Scraper*

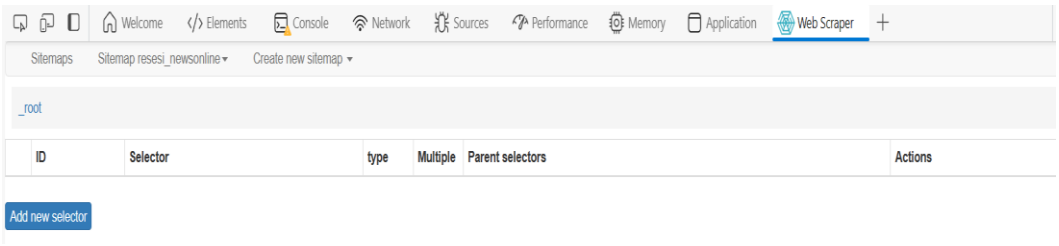
Sitemap gambar 4.1 telah dibuat untuk proses *scraping*, setiap *sitemap* diambil untuk situs berbeda, *sitemap* yang digunakan adalah *sitemap* dengan ID antaranews, resesi_kompas, resesi_cnn, resesi_detik, resesi_idntimes dan sindonews.



Gambar 4. 2 Menambahkan *Sitemap*

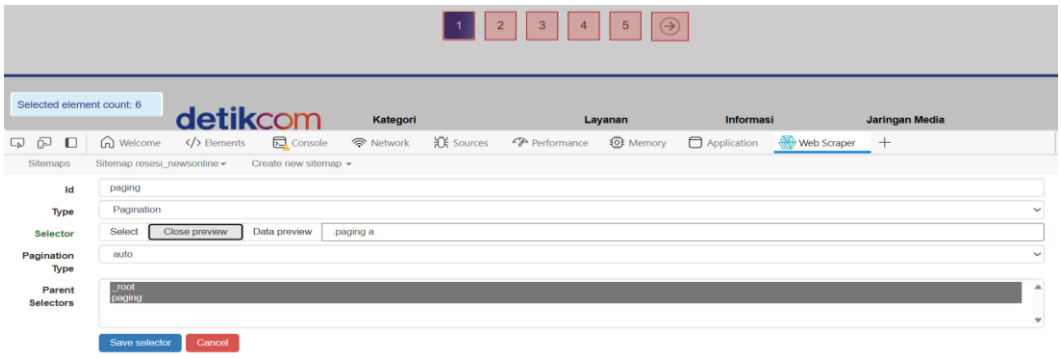
Sitemap name perlu ditulis sesuai kebutuhan. *Start URL* adalah URL utama yang akan menjadi *head* pada *sitemap*, penting untuk memperhatikan dalam mengisi kolom *start URL* dikarenakan *URL* yang didaftarkan akan menjadi

halaman pertama yang di proses oleh web scraper. Halaman *start URL* bukanlah page *home* pada *website* melainkan halaman yang dipilih atau di *scraping*.



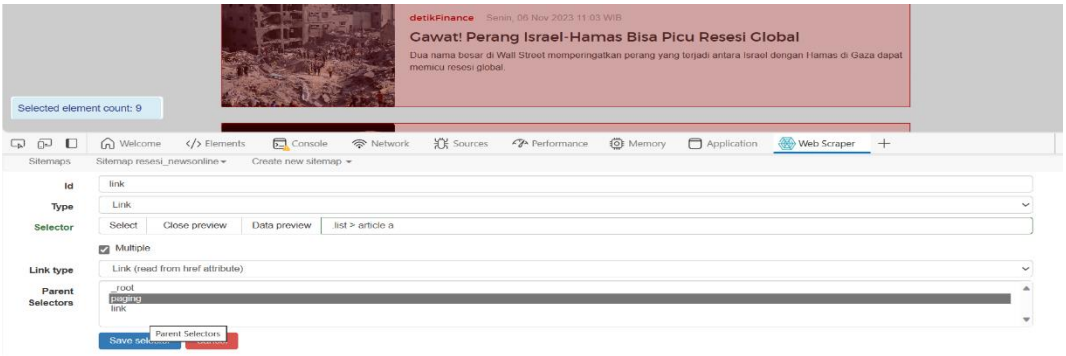
Gambar 4. 3 Tampilan *Selector*

Setelah *sitemap* telah dibuat, gambar 4.3 tambahkan *selector* yang digunakan, *selector* ini bisa digunakan berbagai macam, seperti menambah fitur crawling yang digunakan, melihat tinjauan data yang akan diekstrak dan melihat berbagai fitur yang telah dibuat.



Gambar 4. 4 selector pagination

Pada gambar 4.4 *selector* yang digunakan adalah *pagination*. *Pagination* berfungsi untuk mengganti halaman setelah *sub-link* semua telah diproses. *Pagination* menjadi *selector* pertama karena merupakan *parent* yang akan terus terproses hingga akhir. Pada *selector* tekan pilihan *select* dan pilih bagian *nomor page*, lalu tekan nomor pada *page*, aplikasi dapat automatis memilih *page* lainnya.

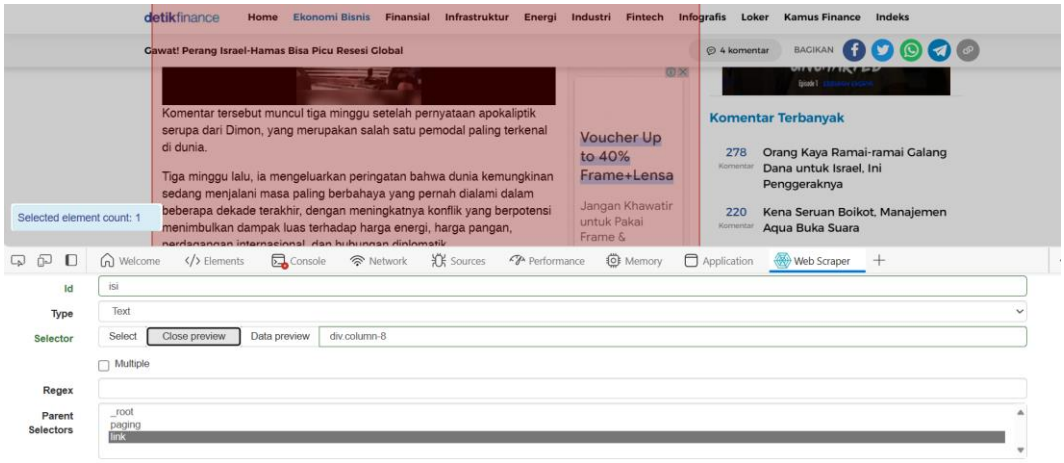


Gambar 4. 5 *Link* Menuju Artikel

Pada gambar 4.5, dibuat *selector* anak dari *paging* yaitu *link*, nantinya mengambil url dari *link* setiap kanal artikel. *Link* bersifat multiple agar semua artikel kanal akan di jelajahi, pada *selector* dapat dilihat data diambil pada *list* dari *tag article:a.* Sama pada proses *pagination*, pada opsi *selector*, *select* dan pilih kolom artikel yang akan di-*crawling.* Aplikasi akan secara otomatis memilih kolom artikel lainnya.

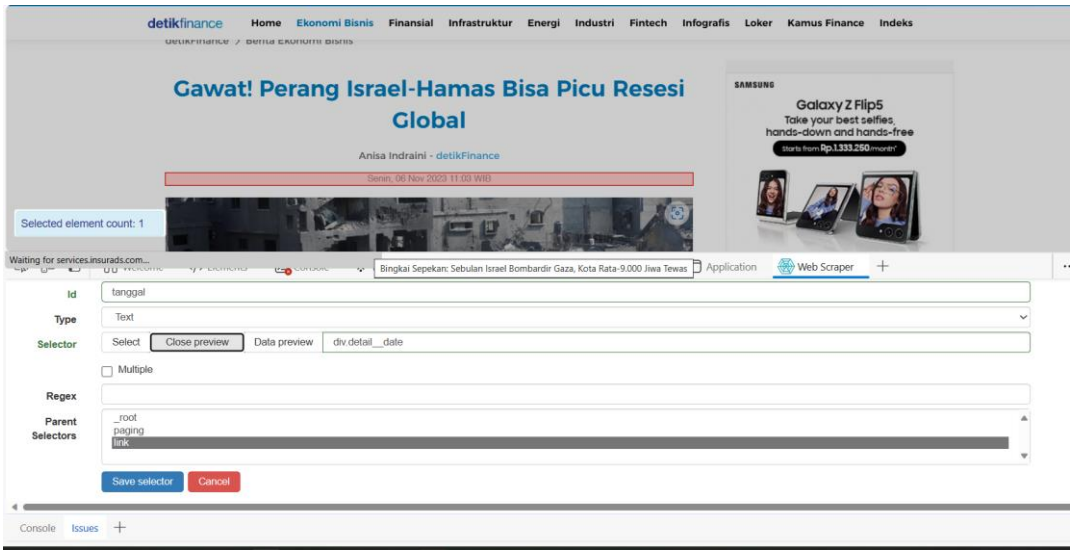
Gambar 4. 6 Selector Teks Judul

Gambar 4.6 merupakan *selector* untuk *scraping* teks berupa judul, pada *selector* dapat dilihat data yang diambil ada data pada *tag h1* pada HTML, judul digunakan untuk membantu memahami isi artikel setelah *scraping.*



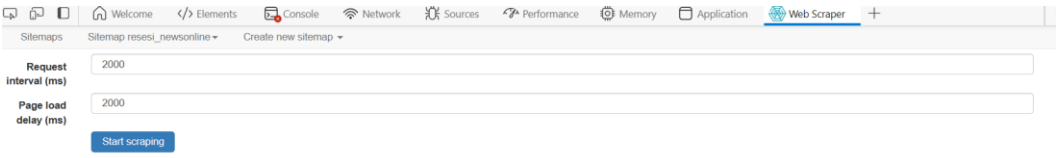
Gambar 4. 7 *Selector* Isi

Selanjutnya pada gambar 4.7, *selector* yang digunakan masih sama yaitu teks, selctor mengambil data content pada artikel untuk digunakan pada pemrosesan model. data yang di ambil adalah data pada *tag div.column-8*.



Gambar 4. 8 *Selector* Tanggal

Selector terakhir adalah *selector* tanggal, pada gambar 4.8, dimana *selector* mengambil tanggal pada *tag div.detail__date*, tanggal diproses dalam membuat data dapat disaring. Data yang diambil dapat diproses hingga tahun 2016-2017 dimana data tersebut berada di luar batas penelitian.



Gambar 4. 9 Proses *Scraping*

Pada gambar 4.9, dilakukan *scraping*, lalu tekan sitemap_resesinews*online* lalu pilih opsi *scrape*, akan ditampilkan waktu yang dibutuhkana setiap *page*-nya, untuk memulai cukup tekan tombol *start scraping*.

Gambar 4. 10 *Export Data Scraping*

Proses diakhiri pada gambar 4.10 dengan melakukan *export data*, data dapat di unduh dengan bertipe excel atau csv, data diunduh dan disimpan komputer pada tempat unduhan.

4.2 *Text Processing*

Tahapan di dalam *text processing* dilakukan dalam beberapa tahapan, yaitu penyaringan tahun, penggabungan data, dan *preprosessing* data.

4.2.1 Penyaringan Tahun

Text processing yang pertama dimulai dari penyaringan tanggal, beberapa dataset yang diambil memiliki distribusi tanggal yang beragam mulai dari 2016-2023. Setiap sumber memiliki tipe format tanggal yang berbeda-beda, sehingga perlu di lakukan penyaringan sebelum dapat digabungkan. Proses penyaringan

menggunakan *regex* untuk mencari pola tanggal berupa tahun (yyyy) teruntuk 2022-2023 saja.

```
fungsi filterTanggal(doc, tanggal):
    docCopy = doc
    hapusNilaiKosong(docCopy)
    untuk setiap baris dalam docCopy:
        tanggal_baru = cariTahun(baris[tanggal])
        baris[tanggal] = tanggal_baru
    hasil = pilihData(docCopy, docCopy[tanggal] > 2021)
    kembalikan hasil
```

Pseudocode diatas akan mengambil parameter doc yang berupa input *dataframe* yang diproses dan tanggal merupakan kolom tanggal pada *dataframe*. Proses dilanjutkan dengan memberikan variabel baru pada *dataframe*. *Dataframe* yang memiliki nilai kosong. Terakhir, data akan di *loop* sebanyak dokumen, diambil tanggal setiap artikel, ambil teks hanya khusus tahun saja (yyyy) dengan regex, lalu disaring dimana data yang diambil hanya data dengan kondisi tahun data diatas 2021. Data yang telah diproses lalu di kembalikan sebagai hasil.

Tabel 4. 2 Jumlah Data Setelah Penyaringan

Sumber Berita	**Sebelum Penyaringan**	**Setelah Penyaringan**
CNN Indonesia	2331	952
Detik.com	1028	480
Kompas.com	313	297
Antaranews	1500	1475
Idntimes	304	304

Data yang telah disaring seperti pada tabel 4.2 memiliki perubahan cukup besar pada berita CNN Indonesia, data sebelum penyaringan berjumlah 2331 dan sesudah penyaringan berjumlah 952 dimana 1379 data tersaring, Detik.com mengalami pengurangan 548 data dari 2331 menjadi 952 data, Kompas.com berkurang 16 data dari 313 menjadi 297 data, dan antaranews berkurang 25 data

dari 1500 data menjadi 1475 data, pengurangan terjadi karena banyak data tidak memenuhi syarat tahun. Idntimes tidak mengalami pengurangan dikarenakan data tersebut sudah dapat di saring langsung pada *website* sehingga tahun pada idntimes sudah memenuhi syarat.

4.2.2 Penggabungan Data

Data yang telah disaring selanjutnya digabungkan, menggunakan fungsi concat pada pandas, *dataframe* digabungkan pada variabel baru untuk memudahkan fungsi preprossing. *Dataframe* yang akan di gabungkan untuk proses selanjutnya hanya *dataframe colomn* “isi” dan “tanggal”.

4.2.3 *Preprosesing Data*

```
fungsi prepros(doc):
    dokumen_baru = []
    untuk setiap dokumen dalam doc:
        lowerCase = lowercasing([dokumen])
        tokenize = tokenizing(lowerCase)
        stopW = sas*Stopword*(tokenize)
        textProcess = removeSpace(stopW)
        stem = stemming(textProcess)
        dokumen_baru.append(stem)
```

Proses selanjutnya adalah *preprosessing* pada umumnya. Data akan diinput pada fungsi prepros yang memanggil fungsi-fungsi lain didalamnya. Fungsi pertama yang akan berjalan adalah *lowercasing*, penggunaannya akan berfungsi di fungsi *stopword*. Kata yang di *lowercase* dapat disaring oleh fungsi *stopword* yang terdiri dari Kumpulan kata yang di dalamnya hanya terdiri huruf kecil, selain itu,

pada proses *lowercasing* juga dilakukan fungsi *cleaning,* dimana kalimat yang memiliki angka, tanda baca, spasi yang terlebih, dan url akan di hapus.

Selanjutnya di fungsi *tokenize* digunakan untuk memecah kalimat menjadi kumpulan kata, pemecahan ini menjadi input pada fungsi *stopword*. Fungsi *stopword* menghapus kata bantu dan kata yang tidak memiliki kontribusi, selain list yang telah disediakan oleh sastrawi, ditambahkan kumpulan kata baru yaitu: dengan, ia, bahwa, oleh, detikcom, widodo, joko, tcopyright, konten, nbaca, idn, times, jakarta, bandung, kompas, kompas.com, konten.

Proses selanjutnya yaitu fungsi *removeSpace* yang akan menghapus spasi dari kata *tokenize* yang telah di saring *stopword,* kata yang memiliki nilai kosong dapat dihilangkan dari *list token*. Proses akhir yaitu *stemming,* mengembalikan kata menjadi kata dasar. Data yang telah diproses akan digabungkan ke dalam *variabel* dokumen baru yang telah dibuat lalu dikembalikan sebagai hasil dari fungsi prepros.

4.3 Membangun *Corpus*

Membangun *corpus* dilakukan untuk mendapat persebaran kata pada *list* data yang telah diproses, dimana tahap ini dilakukan proses TF-IDF dan N-gram. Kedua proses ini dilakukan dengan tujuan untuk mengurangi sekaligus memperkaya kata atau *term* pada *corpus*.

4.3.1 TF-IDF

Gensim telah menyediakan fungsi TF-IDF dari list data dengan memanggil fungsi TfidfModel. TF-IDF pada model digunakan untuk evaluasi *topic modelling* sebagai corpus untuk mendapatkan hasil *coherence score*.

4.3.2 N-gram

Pembuatan fungsi ngram digunakan dengan mengambil data dari hasil preprosseing, dengan jumlah n yang ditentukan, dimana n = 1 adalah unigram, n=2 adalah *bigram*, dan n=3 adalah *trigram*. Penggunaan *bigram* dan *trigram* untuk memperkaya kata semantik untuk mendapatkan informasi lebih bermakna. Pembuatan korpus dibuat untuk melakukan pemetaan angka dengan gram kata. Berikut adalah *pseudocode* yang digunakan dalam pembuatan *trigram*.

```
function make_grams(texts):
    trigram_list = [] # Inisialisasi list untuk menyimpan trigram

    for doc in texts: # Loop melalui setiap dokumen dalam texts
        bigram_doc = apply_bigram_model(doc) # Terapkan model bigram pada dokumen
        trigram_doc = apply_trigram_model(bigram_doc) # Terapkan model trigram pada hasil bigram
        trigram_list.append(trigram_doc) # Tambahkan hasil trigram ke dalam list
```

Pada *pseudocode* diatas, Fungsi akan mengambil daftar teks yang berisi dokumen dan membuat *trigram* dari teks tersebut dengan menerapkan model *bigram* dan *trigram* ke setiap dokumen. Model *bigram* dan *trigram* didapat dari fungsi model pada *library* gensim.

Tabel 4. 3 Jumlah *Corpus* Berdasarkan *N-Gram*

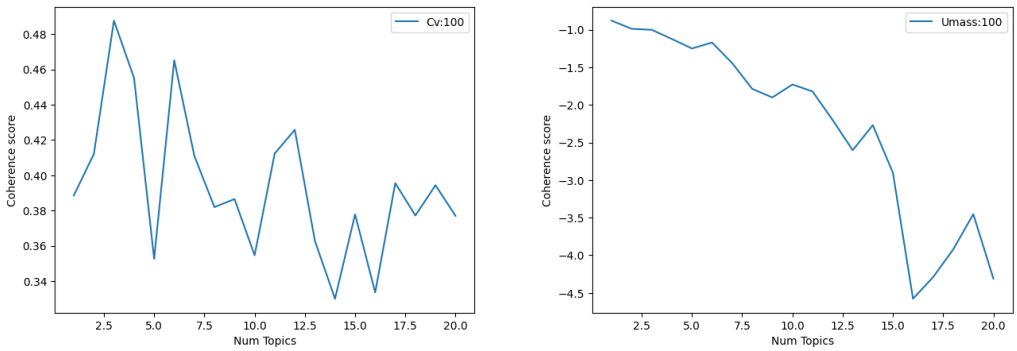
Sebelum ngram	Sesudah ngram
5869	9273

Pada tabel 4.3, jumlah korpus yang sebelumnya masih bersifat unigram atau satu kata saja, berjumlah 5869, seteleah diproses dengan fungsi make_*trigram*s, didapat peningkatan jumlah *corpus*. Peningkatan diakibatkan bertambahnya kata yang telah digabungkan menjadi *bigram* dan *trigram* pada model.

4.4 *Latent Dirichlet Allocation* (LDA)

Pada pemodelan topik dengan metode LDA, variabel bebas yang digunakan sebagai pengujian untuk menentukan model yang digunakan dengan menentukan hasil dari *coherence score*, dimana validasi di lakukan dengan iterasi 100, 500 dan 1000, nilai *coherence score* paling tinggi akan diproses untuk dibandingkan isi topik dengan topik lainnya.

4.4.1 *Coherence Score* Iterasi 100

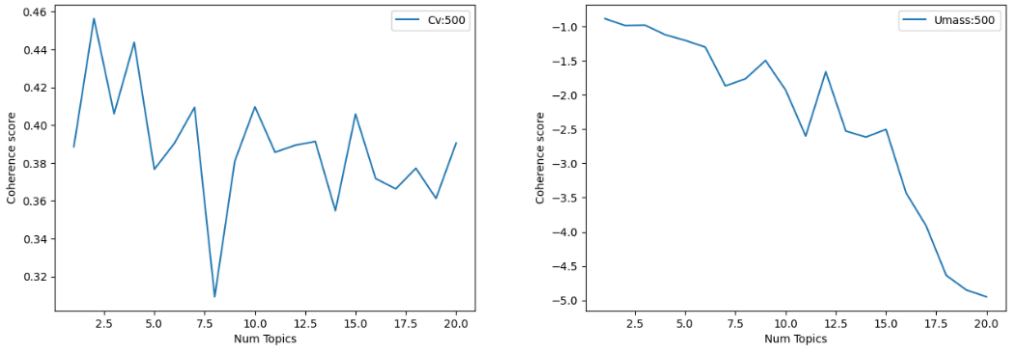


(a) Evaluasi *Cv* 100 (b) Evaluasi *Umass* 100

Gambar 4. 11 Evaluasi *Coherence Score Iteration* 100

Pada gambar 4.11 (a) terlihat evaluasi *Cv* penurunan dan kenaikan ekstrem terlihat dalam grafik, menunjukkan adanya pola garis yang cenderung menuju tren negatif. Evaluasi menggunakan *cohenrence score Cv* menunjukkan ketidakteraturan yang mengindikasikan kualitas model yang kurang baik. *coherence score* menunjukkan performa optimal saat jumlah topik melebihi 3, dengan fluktuasi naik turun. Sementara itu, pada gambar 4.11 (b) evaluasi menggunakan *Umass* menunjukkan fluktuasi yang signifikan di topik 18, mencerminkan model yang awalnya baik namun mengalami fluktuasi dengan dilanjutkan yang menuju tren negatif secara bertahap.

4.4.2 *Coherence score* iterasi 500



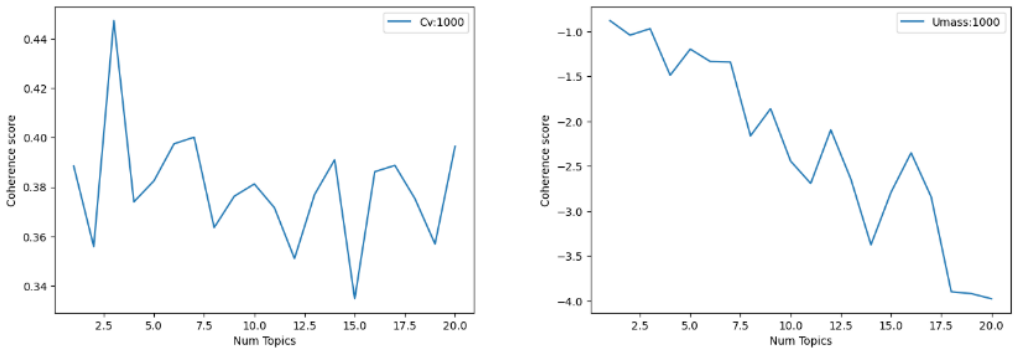
(a) Evaluasi Cv 500 (b) Evaluasi Umass 500

Gambar 4. 12 Grafik *Coherence Score Iteration* 500

Pada gambar 4.12 (a) terlihat evaluasi *Cv* mengalami kenaikan dan dilanjutkan penurunan yang ekstrim, tidak ada tanda-tanda pola grafik akan positif atau negatif. Nilai *coherence score Cv* mulai menunjukkan hasil optimal saat mempertimbangkan dua topik. Namun, saat jumlah topik meningkat di atas dua, skornya cenderung menurun sebelum kembali naik. Ini menunjukkan bahwa pada awalnya, peningkatan jumlah topik dapat mengganggu koherensi, namun

kemudian, setelah titik tertentu, skor koherensi dapat pulih dan trend cendrung menuju positif lagi, evaluasi *Cv* kembali tidak memberi gambaran untuk hasil evaluasi model dikatakan baik. Gambar 4.12 (b) evaluasi pada *Umass* memberikan model yang terus mengalami penurunan walaupun pada titik topik diatas sepuluh mengalami kenaikan, tetapi model Kembali mengalami penurunan. Evaluasi *Umass* memberikan model yang baik dengan topik ke-20 sebagai nilai minimum.

4.4.3 *Coherence Score* Iterasi 1000



(a) Evaluasi *Cv* 1000 (b) Evaluasi *Umass* 1000

Gambar 4. 13 Grafik *Coherence Score Iteration* 1000

Pada gambar 4.13 (a) terlihat evaluasi *Cv* mengalami kenaikan dan penurunan yang ekstrem, tidak ada tanda-tanda pola grafik akan positif atau negatif. Pada analisis koherensi, terlihat bahwa nilai skor koherensi mencapai puncaknya saat jumlah topik melebihi tiga, dengan mencapai nilai tertinggi sebesar. Namun, perlu dicatat bahwa terjadi penurunan drastis setelah titik ini. Pada gambar 4.13 (b) evaluasi *Umass* terlihat adanya kecenderungan negatif secara umum. Meskipun demikian, perlu dicatat bahwa tren ini tidak berjalan secara kontinu atau mulus, karena terdapat beberapa titik di mana terjadi kenaikan dan penurunan nilai secara berulang.

Tabel 4. 4 *Coherence Score* Penentuan Model

Num Topic	**Iterasi 100**		**Iterasi 500**		**Iterasi 1000**	
	Coherence score Cv	***Coherence score Umass***	***Coherence score Cv***	***Coherence score Umass***	***Coherence score Cv***	***Coherence score Umass***
1	0,3885	-0,8798	0,3885	-0,8843	0,3885	-0,8798
2	0,4121	-0,9886	0,4563	-0,9842	0,3559	-1,0413
3	0,4877	-1,0020	0,406	-0,9799	0,4473	-0,9692
4	0,4552	-1,1237	0,4437	-1,1192	0,3740	-1,4878
5	0,3526	-1,2502	0,3766	-1,2022	0,3825	-1,1967
6	0,4652	-1,1715	0,3904	-1,3002	0,3975	-1,3353
7	0,4111	-1,4441	0,4094	-1,8679	0,4001	-1,3410
8	0,3820	-1,7874	0,3092	-1,7632	0,3636	-2,1621
9	0,3865	-1,9016	0,381	-1,4967	0,3763	-1,8605
10	0,3547	-1,7290	0,4097	-1,9269	0,3812	-2,4440
11	0,4124	-1,8217	0,3856	-2,5997	0,3716	-2,6919
12	0,4258	-2,2013	0,3893	-1,6598	0,3511	-2,0960
13	0,3629	-2,6033	0,3913	-2,5266	0,3769	-2,6442
14	0,3300	-2,2700	0,3547	-2,6153	0,3910	-3,3750
15	0,3779	-2,9024	0,4057	-2,5020	0,3348	-2,7904
16	0,3336	-4,5788	0,3717	-3,4303	0,3862	-2,3515
17	0,3956	-4,2914	0,3663	-3,9127	0,3887	-2,8421
18	0,3772	-3,9232	0,3772	-4,6362	0,3753	-3,8975
19	0,3944	-3,4532	0,3612	-4,8487	0,3570	-3,9184
20	0,3770	-4,3108	0,3905	-4,9483	0,3965	-3,9757

Nilai yang ditampilkan pada tabel 4.4 mendapatkan nilai terbaik pada evaluasi *Cv* dengan nilai paling tinggi yaitu 0,4877 dan nilai terbaik kedua dengan nilai 0,4653 dan pada evaluasi *Umass* nilai terbaik di dapat dengan nilai minimum dengan nilai -4,9483 dan nilai paling rendah kedua adalah -4,5788. Terdapat perbedaan hasil yang cukup jauh pada model *Cv* dan *Umass* dalam penentuan topik yang dipilih keperluan analisis. Menggunakan hasil evaluasi menggunakan model pada grafik, model *Umass* memiliki model terbaik dibandingkan *Cv* yang masih tidak memenuhi standar model.

Perbandingan dilakukan terhadap model-model *Umass* berbeda untuk mengevaluasi apakah variasi jumlah iterasi pada korpus mempengaruhi peningkatan hasil *coherence score* dari model yang sedang diuji. Fokus pada

pemilihan topik yang digunakan terpilih dengan menunjukkan nilai *coherence score* terendah pada model *Umass*.

Dari analisis perbandingan model-model, penentuan jumlah topik menggunakan nilai minimum atau paling negatif pada *Umass*, sedangkan untuk membandingkan sebuah koherensi setiap topik dipilih pada nilai yang mendekati 0. Terlihat bahwa model *Umass* dengan 500 iterasi menunjukkan hasil terendah dengan nilai *coherence score* sebesar -4,9483. selanjutnya, model *Umass* dengan 100 iterasi menunjukkan peringkat kedua yang terendah dibandingkan dengan model *Umass* 1000 iterasi yang mencapai nilai paling mendekati koherensi maksimal yaitu 0 dengan nilai -3,9757.

Penulis memutuskan untuk menggunakan hasil evaluasi *coherence score* terendah dan paling mendekati koherensi maksimal 0 dari model *Umass*, terpilih model *Umass* iterasi 1000 sebagai titik fokus untuk analisis lebih lanjut.

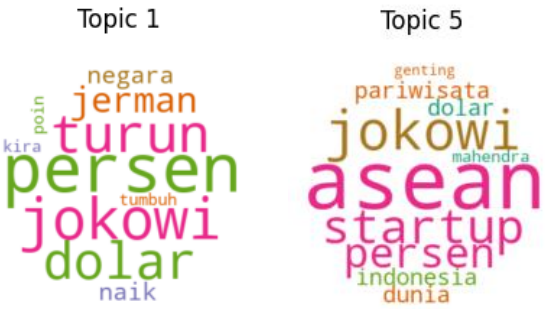
4.5 Analisis Topik

Analisis topik dilakukan dengan memeriksa keterkaitan antara kata pada setiap topik, persebearan kata di dalam topik akan membentuk makna atau wawasan yang berkontribusi pada pemahaman tentang suatu isu. Informasi yang dihasilkan dari analisis semacam ini dapat menjadi sumber bagi penyebab terjadinya resesi, sebagaimana dipaparkan dalam berita *online*.

(a) Topik 0 (b) Topik 2 (c) Topik 4

Gambar 4. 14 *Wordcloud* Topik Kata Membahas Harga Minyak

Gambar 4.14 menampilkan persebaran kata topik yang membahas kata minyak sebagai kata dominan, persebaran kata minyak dihubungkan dengan harga minyak dan dampak pada dunia internasional, hal ini karena dampak dari pembatasan harga minyak di Rusia dijelaskan pada topik 2. Dilanjutkan dengan kata dolar yang merupakan nilai mata uang yang digunakan secara internasional sebagai tranasaksi jual beli mengalami penurunan imbas dari harga minyak. Kata China pada gambar 4.14 (a) mengenai topik 0 merujuk pada salah satu negara yang terdampak selama penurunan nilai dolar, kata lain pada topik 0 seperti moeldoko, ia adalah kepala staf kepresidenan yang menjelaskan kondisi dagang di Indonesia. Gambar 4.14 (b) dan (c) yaitu topik 2 dan topik 4 terdapat pembahasan seperti dagang, reksa dana dan saham yang membahas kondisi perdagangan saham dan investasi.



(a) Topik 1 (b) Topik 5

Gambar 4. 15 *Wordcloud* Topik Kata Membahas Jokowi dan Asean

Gambar 4.15 memiliki kemunculan kata terbanyak berbeda yaitu persen dan ASEAN, dengan mengecualaikan kata terbanyak, kata seperti Presiden Jokowi yang merupakan presiden Republik Indonesia menjadi kata terbanyak kedua muncul, Presiden Jokowi pada kedua topik membahas tentang Kebijakan ekonomi Presiden Jokowi untuk negara Indonesia mengalami pertumbuhan terutama di bidang *startup* dan pariwisata terutama di ASEAN dan di dunia. Hal lain yang dibahas pada topik 1 terlihat pada gambar 4.15 (a) berputar pada pembahasan perekonomian negara jerman yang turun karena permasalahan bahan bakar sehingga menaikan peluang resesinya, selain itu, terlihat pada gambar 4.15 (b) dimana topik 5 membahas dengan fokus kepada sektor industri *startup* dan pariwisata mampu mempertahankan Indonesia dari keadaan genting yang terjadi di dunia.

(a) Topik 3 (b) Topik 7 (c) Topik 12

Gambar 4. 16 *Wordcloud* Kata Membahas Ekonomi Amerika dan Eropa

Gambar 4.16 menampilkan topik seputar hal yang berkaitan dengan perekonomian negara barat terutama Amerika dan Eropa, hal ini dilihat dari gambar 4.16 (a) dimana kata terbesar pada topik 3 yaitu *Silicon Valley Bank* (SVB) yang merupakan salah satu bank terbesar di Amerika Serikat mengalami keruntuhan, hal ini mengakibatkan peluang resesi di Amerika semakin besar. Gambar 4.16 (a) juga membahas beberapa hal lain yang terkandung pada topik 3 adalah mengenai Grab, Grab adalah perusahaan teknologi yang berkembang selama pandemi, kata "ruf" merupakan nama dari Wakil Presiden Republik Indonesia yaitu Ma'ruf Amin yang berusaha memajukan ekonomi Indonesia dengan kreativitas dan Kerjasama dengan Perusahaan seperti Grab.

Pada gambar 4.16 (b) terlihat topik 7 membahas mengenai upaya lembaga Bank Sentral Amerika (FED) berusaha menekan laju pertumbuhan inflasi di negaranya dengan menarik investor menutup permasalahan pada bank yang tutup, selain FED, eropa sendiri terkena dampak dari peristiwa-peristiwa topik sebelumnya seperti negara Swedia yang menjadi salah satu *topwords* ditampilkan pada gambar 4.16 (c) di topik 12 dimana membahas pertumbuhan yang hanya naik

satu persen dan diperikirakan akan mengalami inflasi, hal ini disebabkan Swedia bergantung dengan bisnis dengan negara lain.

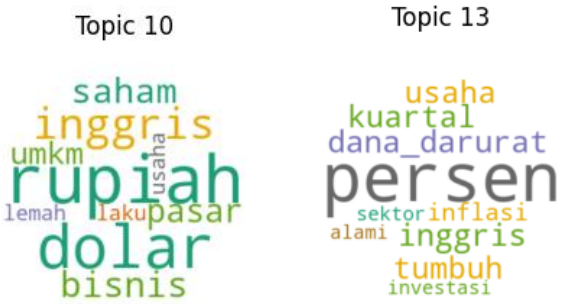
(a) Topik 6 (b) Topik 9 (c) Topik 11

Gambar 4. 17 *Wordcloud* Topik Kata Membahas Mata Uang Dolar dan Global, Deflasi dan Negara Inflasi

Gambar 4.17 menjelaskan keadaan ekonomi di Asia, dampak dan kepanikan yang diakibatkan dari bank sentral Amerika, mengakibatkan mata uang Yen Jepang juga ikut turun, di lain pihak, China berhasil menguatkan perekonomian negaranya. Upaya negara Jepang untuk menaikkan upah untuk menghilangkan resiko deflasi, di lain hal, kedua topik menjelaskan upaya Indonesia dalam menghadapi tekanan global, terlihat pada gambar 4.17 (a) pada topik 6 dikembangkan usaha UMKM, sedangkan di gambar 4.17 (b) pada topik 9 deflasi terjadi di Indonesia diikuti dengan pertumbuhan nilai ekspor, beberapa hal lain yang dibahas pada topik 9 adalah China, Korsel dan Elon Musk yang diusahakan untuk menjadi investor dan membangun pabrik-pabrik di Indonesia. Gambar 4.17 (c) topik 11 sendiri merupakan keadaan ekonomi Indonesia selama terjadi inflasi global, dimana kondisi ekonomi negara Indonesia mengalami pertumbuhan, hal ini disampaikan juga oleh Rukita selaku Perusahaan *startup* yang menyampaikan kesiapan menghadapi inflasi ekonomi global.

Gambar 4. 18 *Wordcloud* Topik 8 Kata Membahas Tokoh Ekonomi

Gambar 4.18 banyak menjelaskan Ariston Tjendra sebagai pengamat pasar uang menjelaskan kekuatan nilai rupiah terhadap mata uang dolar ketika sedang inflasi, Ariston juga menjelaskan kekuatan nilai rupiah dibayangi dengan isu bahan bakar minyak (BBM), selain itu, Ariston juga menjelaskan naiknya harga emas dikarenakan minat investasi emas menjadi minat di Indonesia. Dari beberapa kata pada topik, terdapat kata "Purbaya" adalah Ketua Dewan Komisioner Lembaga Penjamin Simpanan (LPS), purbaya juga menjelaskan keadaan ekonomi Indonesia menyikapi keadaan ekonomi global.



(a) Topik 10 (b) Topik 13

Gambar 4. 19 *Wordcloud* Topik Kata Membahas Rupiah dan Inggris

Topik 10 dan topik 13 terlihat di gambar 4.19 memiliki kata "Inggris", dimana permasalahan dialami negara Inggris pada sektor perekonomian yang melemah karena mengalami inflasi, tetapi, kata Inggris tidak terkait dengan kata-

kata lainnya pada topik masing-masing. Hal ini ditampilkan pada gambar 4.19 (a) dengan topik 10 yang lebih membahas tentang pelaku bisnis pasar saham dan usaha UMKM, lain hal pada gambar 4.19 (b) di topik 13 yang lebih ke pembahasan mengenai pengadaan dana darurat imbas dari kondisi penurunan ekonomi global.

Gambar 4. 20 *Wordcloud* Topik 14 Kata Membahas Harga Dolar

Gambar 4.20 merupakan persebaran dari topik 14, dimana kata dolar merupakan kata dominan dilanjutkan dengan emas dan rupiah ketiga hal tersebut berkaitan dengan perdagangan *foreign exchange* atau Forex. Topik diatas mengaitkan kondisi keuangan Amerika yang melemah dan memasuki periode depresiasi terhadap mata uang asing selain itu, harga emas mengalami penguatan. Depresiasi mata uang adalah ketika uang kita menjadi lebih sedikit nilainya dibandingkan sebelumnya. Terkahir terdapat kata Jokowi yang tidak memiliki hubungan dengan kata lainnya pada topik tersebut.

Gambar 4. 21 *Wordcloud* Topik 16 Kata Membahas Disiplin, Makassar dan hal lain

Persebaran kata pada topik 16 dapat dilihat pada gambar 4.21, kata displin terhubung dengan kata Prita yang merupakan seorang konsultan keuangan, ia menjelaskan untuk displin dalam mengelola keuangan. Kata lain pada topik seperti pintu, project, dan industri furnitur berkaitan dengan produk industry Indonesia yang banyak di ekspor ke luar negeri. Kata enablr pada topik merujuk pada Perusahaan *start up ecommerce* yang tidak akan melakukan PHK terhadap karyawan dan berfokus meningkatkan partner, selain itu, kata jerman pada topik tidak merujuk pada apapun dalam topik.

Gambar 4. 22 *Wordcloud* Topik 17 Kata Membahas Perekonomian Sumatera Utara

Persebaran kata pada topik 17 ditampilkan pada gambar 4.22, kata dengan frekuensi kemunculan terbesar adalah Sumut, Sumut banyak membahas mengenai pertumbuhan ekonomi di provinsi Sumatera Utara dan kekuatan UMKM dalam meningkatkan perekonomian Sumut, lain hal dengan kata KPR yang menjelaskan tentang keuntungan menggunakan KPR dalam membeli aset rumah dan kata koperasi yang lebih ke pentingnya koperasi sebagai penopang ekonomi menghadapi resesi global. Kata lainnya seperti Fed Fund Rate (FFR) yang menaikkan suku bunga banknya, kata India Brazil merupakan negara yang tergabung pada BRICH dan juga negara yang siap menghadapi inflasi global. Kata lainnya seperti Vietnam

merupakan negara yang juga bekerja cukup baik ditengah ketidak pastian global. Terdapat satu kata yaitu malaria yang sama sekali tidak terkait dalam topik.

(a) Topik 15 (b) Topik 18 (c) Topik 19

Gambar 4. 23 *Wordcloud* Topik Kata Membahas PHK dan Peran UMKM

Gambar 4.23 menampilkan 3 topik yang memiliki kedekatan dalam pembahasannya, dimana topik 15 dan topik 18 membahas PHK. Pada gambar 4.23 (a) dijelaskan pada topik 15 bahwa banyak karyawan terkena PHK, untuk mengatsinya, usaha UMKM menjadi pondasi mengatasi permasalahan PHK, UMKM dimudahkan dalam aktivitas agar semakin tumbuh, hal ini dengan bantuan perbankan yang memberikat saluran kredit kepada pelaku usaha UMKM untuk mengembangkan usaha mereka. Kondisi PHK juga menjadi permasalahan terlihat pada gambar 4.23 (b) dimana topik 18 lebih membahas ke pendapat tokoh Said Iqbal sebagai ketua partai buruh meminta untuk kenaikan upah UMP, menolak PHK dan normalisasi harga pangan. Said Iqbal juga mengatakan bahwa buruh juga harus siap menghadapi inflasi yang telah terjadi di negara-negara Eropa.

Gambar 4.23 (c) topik 19 menjelaskan mengenai poin sebagai kata dominan, poin disini mengacu dalam saham dan IHSG yang juga berlaku sama pada kata persen. Terdapat Goto yang merupakan perusahaan *start up* yang juga melakukan PHK karyawannya imbas dari kondisi saham yang terus menuju negatif

dan imbas dari inflasi global. Pembahsan lain yang muncul pada topik 19 adalah kata usaha UMKM yang juga telah dibahas topik 15, lalu terdapat kata sekolah yang masih tidak diketahui hubungannya dengan topik lalu kata Jabar yang merujuk pada pemerintahan Jawa Barat yang menyiapkan bantuan kepada pekerja yang terkena dampak PHK.

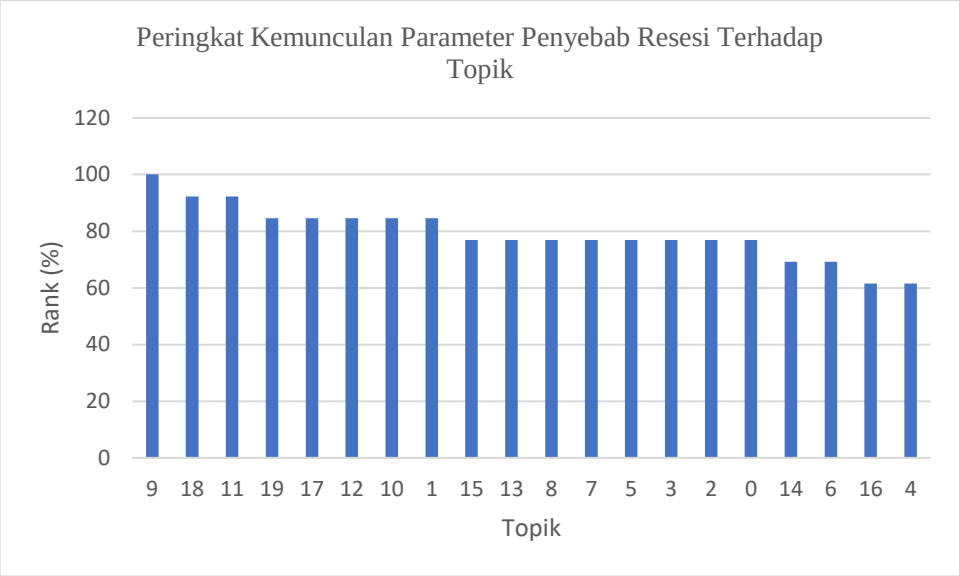
Tabel 4. 5 Interpretasi Persebaran Kata

Topik	Persebaran kata	Interpretasi
0	minyak, saham, china, persen, turun, harga_minyak, dagang, dollar, moeldoko, naik	Pengaruh harga minyak terhadap perdagangan dan ekonomi global
1	persen, jokowi, turun, dolar, jerman, negara, naik, tumbuh, kira, poin	Kebijakan Ekonomi Presiden Jokowi dan Pertumbuhan ekonomi Negara Jerman turun
2	minyak, dolar_per_barel, rusia, persen, dolar, dagang, harga_minyak, turun, reksa_dana, pasar	Dampak pembatasan harga minyak di Rusia dan pergerakan dolar terhadap Pasar dan kondisi perdagangan saham dan investasi
3	svb, dolar, persen, grab, saham, ruf, poin, seks, kreativitas, naik	Dampak Sillicon Valley Bank yang bankrut terhadap saham dan dolar
4	minyak, dolar, persen, turun, saham, naik, harga_minyak, emas, harga, pasar	Dinamika Harga Minyak dan Emas dalam Kaitannya dengan Pergerakan Saham, Dolar, dan kondisi perdagangan saham dan investasi
5	asean, jokowi, startup, persen, pariwisata, indonesia, dolar, dunia, mahendra, genting	Pengaruh Kebijakan Pariwisata, Startup, dan Hubungan ASEAN dalam Konteks Kepemimpinan Jokowi
6	dolar, yen, persen, naik, umkm, bulan, turun, usaha, china, fed	perubahan nilai tukar Dolar berdampak pada nilai Yen, UMKM di Indonesia meyikapi kebijakan moneter The Fed dan investor bisnis dari China
7	persen, negara, the_fed, kerja, investasi, naik, tumbuh, inflasi, investor, usaha	Pertumbuhan Investasi, Kebijakan The Fed,untuk menambah investor untuk mengurangi dampak Inflasi, dan Kondisi Usaha di negara Amerika.
8	ariston, rupiah, dolar, persen, per_dolar, inflasi, purbaya, minyak, gram_harga_emas, kuat	Perubahan Nilai Tukar Rupiah terhadap Dolar dan Faktor-faktor Ekonomi seperti Inflasi, Harga Minyak, dan Emas dalam Konteks Ariston dan Purbaya
9	deflasi, persen, korsel, ekspor, tumbuh, indonesia, china, elon_musk, harga, disney	Pertumbuhan Ekspor Indonesia dan masuknya investor dari Korsel dan China di Tengah Kondisi Deflasi
10	rupiah, dolar, inggris, bisnis, saham, pasar, umkm, lemah, usaha, laku	fluktuasi nilai tukar Rupiah-Dolar terhadap bisnis, saham, negara Inggris ekonomi melemah karena inflasi dan performa UMKM di pasar, khususnya saat Rupiah melemah
11	negara, tumbuh, inflasi, indonesia, rukita, kondisi, persen, ekonomi, dollar, rusia	*Startup* Indonesia siap menghadapi ancaman inflasi ekonomi global
12	persen, bank, inflasi, swedia, tumbuh, proyeksi, kira, negara, naik, bisnis	sektor perbankan dan bisnis, serta perhatian pada inflasi dan kebijakan bank global juga

Topik	Persebaran kata	Interpretasi
		berdampak pada proyeksi pertumbuhan ekonomi Swedia,
13	persen, dana_darurat, inggris, usaha, tumbuh, kuartal, inflasi, investasi, sektor, alami	kondisi pertumbuhan ekonomi global terdampak inflasi, berdampak pada negara Inggris ekonomi melemah. sektor-sektor investasi, ekonomi yang berbeda, kemungkinan penggunaan dana darurat dalam situasi keuangan yang sulit
14	dolar, emas, rupiah, dolar_per_ounce, jokowi, persen, ibrahim, kuat, lemah, depresi	Evaluasi kondisi ekonomi Indonesia saat terjadi depresiasi nilai tukar mata uang dan harga emas
15	umkm, usaha, phk, perban, kredit, salur_kredit, karyawan, tumbuh, laku, mampu	UMKM menjadi penopang dengan pemberian akses kredit, perlindungan karyawan PHK, dan pertumbuhan bisnis.
16	disiplin, pintu, makassar, persen, enablr, prita, jerman, partner, project, industri_furnitur	Disiplin *start up* Indonesia dalam membuat proyek terutama ekspor industry furnitur.
17	sumut, kpr, persen, koperasi, ffr, vietnam, cadang_devisa, malaria, negara, india_brazil	Ekonomi Sumatra Utara dalam menyikapi inflasi dengan kekuatan UMKM dan koperasi, keuntungan KPR, dan negara yang siap menghadapi inflasi global
18	buruh, phk, negara, said_iqbal, kerja, usaha, pangan, eropa, indonesia, upah	Seruan serikat buruh untuk menghentikan PHK, menaikkan upah, dan normalisasi harga pangan imbas keadaan di Eropa
19	poin, ihsg, goto, umkm, usaha, persen, jabar, sekolah, eropa, saham	perubahan nilai IHSG pengaruh saham GOTO, dan upaya pemerintahan jabar dalam kegiatan UMKM untuk menghadapi resesi global

Persebaran kata dapat dengan interpretasi ditampilkan pada tabel 4.5, hubungan kata dengan interpretasi berdasarkan peristiwa yang terjadi, Hasil dari interpretasi menyampaikan bahwa dominasi interpretasi dari persebaran dua puluh topik dengan sepuluh kata dengan probabilitas terbesar. Interpretasi menunjukkan adanya korelasi dengan hal yang menjadi penyebab resesi berdasarkan penjelasan pada bab 1 yang diperoleh parameter yang bisa menjadi penyebab resesi yaitu kata ekonomi, negatif, fluktuasi, inflasi, produk domestik bruto, produksi, distribusi, konsumsi, investasi, phk, krisis, turun, dan deflasi. Dalam mengetahui apakah sebuah topik memiliki parameter penyebab resesi dapat dengan menggunakan

rumus 3.8 dimana diperoleh bahasan mengenai parameter resesi dengan pada setiap topik.



Gambar 4. 24 Peringkat Kemunculan Parameter Resesi Terhadap Topik

Terlihat pada gambar 4.24 bahwa semua topik ada dan memberikan kata penyebab resesi. Namun, gambar 4.24 hanya mempresentasikan seberapa banyak topik ini membahas parameter penyebab resesi. Bila dihubungkan dengan interpretasi pada tabel 4.5, banyak kata parameter pada topik lebih condong dalam hal yang perlu diwaspadai dan merupakan hal yang mengancam ekonomi.

Bila dihubungkan parameter penyebab resesi dengan tabel 4.5 mengenai interpretasi dari persebaran dua puluh topik dengan sepuluh kata, korelasi menunjukkan beberapa faktor yang menjadi penyebab dari resesi dimana pada parameter ekonomi yang negatif atau turun yang diakbatkan oleh harga minyak (topik 0) dan embargo minyak oleh rusia (topik 2), mengakibatkan harga investasi seperti saham turun (topik 4).

Merujuk parameter penyebab resesi yaitu kondisi investasi juga dibahas seperti pada perubahan nilai tukar rupiah yang terus berubah dipengaruhi nilai dollar (topik 3, topik 12) dan inflasi (topik 8) untuk mengahadapi hal tersebut perlu dilakukan evaluasi ekonomi indonesia terhadap nilai tukar uang dan harga emas (topik 14) yang secara tidak langsung akan memepengaruhi pasar saham Indonesia atau IHSG (topik 19). Kondisi investasi mata uang asing (dollar) membuat melemahnya beberapa sektor seperti sektor UMKM (topik 6, topik10).

Beberapa isu berita lebih menunjukan berapa banyak topik yang berkaitan dengan resesi namun ada beberapa topik yang menunjukan perlawanan atau hal yang perlu disiapkan dalam resesi seperti perkembangan Kebijakan-kebijakan oleh presiden joko widodo (topik 1) yang mengakibatkan kebijakan kebijakan baru di bidang parawisata dan starts up, tidak lupa hubungan bileteral di tingkat ASEAN (topik 5). Indonesia juga memiliki UMKM dan startup yang telah bersiap menghadapi resesi (topik 16, topik 9, topik 11). dimana UMKM menjadi penopang penyerapan pekerja-pekerja PHK (topik 15) bisa diambil kesimpulan bahwa ada beberapa potensi resesi yang bisa terjadi di indonesia namun beberapa berita menunjukan kesiapan indonesia menghadapi resesi, terutama pada bidang UMKM.

4.6 Integrasi Islam

Penelitian ini mengintegrasikan amalan-amalan pentingnya untuk *tabayyun* yaitu mencari kejelasan atau kebenaran terhadap berita yang beredar. Pentingnya mendapatkan pencerahan atau informasi sebaik mungkin agar tidak menjadi bahaya terhadap diri sendiri dan orang lain. *Tabayyun* dijelaskan pada al-qur’an Q.S Al-Hujurat ayat 6 yang berbunyi:

يَا أَيُّهَا الَّذِينَ آمَنُوا إِنْ جَاءَكُمْ فَاسِقٌ بِنَبَإٍ فَتَبَيَّنُوا أَنْ تُصِيبُوا قَوْمًا بِجَهَالَةٍ فَتُصْبِحُوا عَلَىٰ مَا فَعَلْتُمْ نَادِمِينَ

"Hai orang-orang yang beriman, jika datang kepadamu orang fasik membawa suatu berita, maka periksalah dengan teliti agar kamu tidak menimpakan suatu musibah kepada suatu kaum tanpa mengetahui keadaannya yang menyebabkan kamu menyesal atas perbuatanmu itu." (Q.S Al-Hujurat: 6)

Dijelaskan oleh tafsir al-jalalayn, (Hai orang-orang yang beriman! Jika datang kepada kalian orang fasik membawa suatu berita) (maka periksalah oleh kalian) kebenaran beritanya itu, apakah ia benar atau berdusta. Menurut suatu qiraat dibaca Fatatsabbatuu berasal dari lafal Ats-Tsabaat, artinya telitilah terlebih dahulu kebenarannya (agar kalian tidak menimpakan musibah kepada suatu kaum) menjadi Maf'ul dari lafal Fatabayyanuu, yakni dikhawatirkan hal tersebut akan menimpa musibah kepada suatu kaum (tanpa mengetahui keadaannya) menjadi Hal atau kata keterangan keadaan dari Fa'il, yakni tanpa sepengetahuannya (yang menyebabkan kalian) membuat kalian (atas perbuatan kalian itu) yakni berbuat kekeliruan terhadap kaum tersebut (menyesal) selanjutnya Rasulullah Shallallahu`alaihi Wa Sallam. mengutus Khalid kepada mereka sesudah mereka kembali ke negerinya. Ternyata Khalid tiada menjumpai mereka melainkan hanya ketaatan dan kebaikan belaka, lalu ia menceritakan hal tersebut kepada Rasulullah Shallallahu`alaihi Wa Sallam. Ayat diatas menjelaskan bahaya berdusta dan pentingnya mencari kebenaran terlebih dahulu. Perlunya dilakukan klaririfikasi terhadap suatu berita agar tidak memberikan informasi yang salah atau tidak benar

BAB V
KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil pembahasan dari penelitian yang telah dilakukan, kesimpulan yang dapat diambil adalah: pemodelan topik dengan metode *Latent Dirichlet Allocation* (LDA) menggunakan data dari media *online* sebanyak 5476 data artikel. Selanjutnya data artikel dilakukan proses *text-processing* yang memiliki tahap *filtering*, *lowercasing*, *tokenzing*, *stopword*, dan *removespace* dilanjutkan dengan tahap *n-gram* sebelum pembobotan *Term frequency invers document frequency* (TF-IDF). Tahap mapping dari string menjadi angka yang kemudian disimpan dalam *bag of word*. Selanjutnya diproses sebagai dictionary dan corpus untuk dijadikan model latih LDA dan divalidasi dengan menentukan *Coherence score*. Didapat model iterasi 1000 sebagai model terbaik dengan metode *Umass* dengan coherence score mencapai -3,9757. Model *Umass* terpilih karena model *Cv* tidak menunjukkan grafik yang baik dibandingkan model *Umass*, selanjutnya pada model *Umass*, dipilih topik sebanyak 20 sebagai nilai dengan *coherence score* paling minimum, model *Umass* iterasi 1000 memiliki nilai terbaik di antara ketiga iterasi karena memiliki nilai paling mendekati 0 atau mendekati koheransi maksimal. Nilai tersebut digunakan untuk penentuan jumlah topik pada metode LDA yaitu sebanyak 20 topik. Dari analisis 20 topik dapat diperoleh kesimpulan beberapa parameter yang menjadi penyebab resesi di Indonesia adalah ekonomi negatif, penurunan investasi , dan inflasi ekonomi.

5.2 Saran

Penulis menyarankan untuk penelitian kedepannya sehingga bisa dikembangkan lebih lanjut:

1. Dataset diperbanyak agar hasil validasi model semakin baik dan persebaran topik semakin bervariasi.
2. Kata *stopword* diperbanyak karena banyak kata yang tidak memiliki makna penting dan tidak berkontribusi pada topik ikut ambil bagian dalam topik.
3. Parameter validasi bisa ditambah atau menggunakan validasi lainnya seperti evaluasi oleh manusia.

DAFTAR PUSTAKA

Alfanzar, A. I., Khalid, K., & Rozas, I. S. (2020). *Topic modelling* Skripsi Menggunakan Metode Latent Diriclhet Allocation. *JSiI (Jurnal Sistem Informasi)*, *7*(1), 7. https://doi.org/10.30656/jsii.v7i1.2036

Al-khairi, Y. U., Wibisono, Y., & Putro, B. L. (2017). Deteksi Topik Fashion Pada Twitter Dengan Latent Dirichlet Allocation. *Jurnal Aplikasi Dan Teori Ilmu Komputer*, *1*(1), 1–10.

Anugrah, I. G. (2021). Penerapan Metode *N-gram* dan Cosine Similarity Dalam Pencarian Pada Repositori Artikel Jurnal Publikasi. *Building of Informatics, Technology and Science (BITS)*, *3*(3), 275–284. https://doi.org/10.47065/bits.v3i3.1058

Blandina, S., Noor Fitrian, A., & Septiyani, W. (2020). Strategi Menghindarkan Indonesia dari Ancaman Resesi Ekonomi di Masa Pandemi. *Efektor*, *7*(2), 181–190. https://doi.org/10.29407/e.v7i2.15043

Chairullah, N. (2020). *Topic modelling* pada Sentimen terhadap Headline Berita *Online* Berbahasa Indonesia. *Universitas Islam Indonesia*, *2769*.

Firdaus, M. R., Rizki, F. M., Gaus, F. M., & Susanto, I. K. (2020). Analisis Sentimen Dan *Topic modelling* Dalam Aplikasi Ruangguru. *J-SAKTI (Jurnal Sains Komputer Dan Informatika)*, *4*(1), 66. https://doi.org/10.30645/j-sakti.v4i1.188

Ghani, N. A., Hamid, S., Targio Hashem, I. A., & Ahmed, E. (2019). Sosial media *big data* analytics: A survey. *Computers in Human Behavior*, *101*(August), 417–428. https://doi.org/10.1016/j.chb.2018.08.039

KBBI. (2023). *resesi*. 2023. https://kbbi.kemdikbud.go.id/entri/resesi

Martínez-Castaño, R., Pichel, J. C., & Losada, D. E. (2020). A *big data* platform for real time analysis of signs of depression in sosial media. *International Journal of Environmental Research and Public Health*, *17*(13), 1–23. https://doi.org/10.3390/ijerph17134752

Maulana, A., & Fajrin, A. A. (2018). Penerapan *Data mining* Untuk Analisis Pola Pembelian Konsumen Dengan Algoritma Fp-Growth Pada Data Transaksi Penjualan Spare Part Motor. *Klik - Kumpulan Jurnal Ilmu Komputer*, *5*(1), 27. https://doi.org/10.20527/klik.v5i1.100

Moodley, A., & Marivate, V. (2019). *Topic modelling* of news articles for two consecutive elections in South Africa. *2019 6th International Conference on Soft Computing and Machine Intelligence, ISCMI 2019*, 131–136. https://doi.org/10.1109/ISCMI47871.2019.9004342

Nastuti, A. (2019). Amelia Nastuti 1) , Syaiful Zuhri Harahap 2). *Teknik Data mining Untuk Penentuan Paket Hemat Sembako Dan Kebutuhan Harian Dengan Menggunakan Algoritma Fp-Growth*, *7*(3), 111–119.

Naury, C., Fudholi, D. H., & Hidayatullah, A. F. (2021). *Topic modelling* pada Sentimen Terhadap Headline Berita *Online* Berbahasa Indonesia Menggunakan LDA dan LSTM. *Jurnal Media Informatika Budidarma*, *5*(1), 24. https://doi.org/10.30865/mib.v5i1.2556

Niu, Y., Ying, L., Yang, J., Bao, M., & Sivaparthipan, C. B. (2021). Organizational business intelligence and decision making using *big data* analytics. *Information Processing and Management*, *58*(6), 102725. https://doi.org/10.1016/j.ipm.2021.102725

Nugraha, A. K., Dewi, Y. P., Informatika, T., Informasi, F. T., Luhur, U. B., Informasi, S., Informasi, F. T., Luhur, U. B., & Utara, P. (2017). Evaluasi Efektivitas Pencarian Dokumen Html Pada Penerapan Stemmed Term Vector Model Dengan Pembobotan Logaritma Frekuensi Term. *Evaluasi Efektivitas Pencarian Dokumen Html Pada Penerapan Stemmed Term Vector Model Dengan Pembobotan Logaritma Frekuensi Term*, *1*, 11.

Nurlayli, A., & Nasichuddin, Moch. A. (2019). Topik Modelling Penelitian Dosen Jptei Uny Pada Google Scholar Menggunakan Latent Dirichlet Allocation. *Elinvo (Electronics, Informatics, and Vocational Education)*, *4*(2), 154–161. https://doi.org/10.21831/elinvo.v4i2.28254

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*. 26–27.

Sikumbang, E. D. (2018). Penerapan *Data mining* Penjualan Sepatu Menggunakan Metode Algoritma Apriori. *Jurnal Teknik Komputer AMIK BSI (JTK)*, *Vol 4*, *No.*(September), 1–4.

Togatorop, P. R., Simanjuntak, R. P., Manurung, S. B., & Silalahi, M. C. (2021). Pembangkit Entity Relationship Diagram Dari Spesifikasi Kebutuhan Menggunakan Natural Language Processing Untuk Bahasa Indonesia. *Jurnal Komputer Dan Informatika*, *9*(2), 196–206. https://doi.org/10.35508/jicon.v9i2.5051

Asrirawan, A., Permata, S. U., & Fauzan, M. I. (2022). Pendekatan Univariate Time Series Modelling untuk Prediksi Kuartalan Pertumbuhan Ekonomi Indonesia Pasca Vaksinasi COVID-19. *Jambura Journal of Mathematics*, *4*(1), 86–103. https://doi.org/10.34312/jjom.v4i1.11717

Blandina, S., Noor Fitrian, A., & Septiyani, W. (2020). Strategi Menghindarkan Indonesia dari Ancaman Resesi Ekonomi di Masa Pandemi. *Efektor*, *7*(2), 181–190. https://doi.org/10.29407/e.v7i2.15043

Gumelar, F., Fathia Luthfiah Nur Solihat, Ni Gusti Ayu Putu Meyrasinta Susila, &

Resa Septiani Pontoh. (2023). Ancaman Resesi: Peran UMKM Dalam Akselerasi Perekonomian Jawa Barat Pasca Pandemi. *Emerging Statistics and Data Science Journal*, *1*(2), 291–300. https://doi.org/10.20885/esds.vol1.iss.2.art29

Jacob, J., & Waibot, Z. (2022). Mengukur Output Gap Ekonomi Maluku Utara (Pendekatan Hodrick-Prescott Filter). *Jurnal Ekonomi Dan Statistik Indonesia*, *2*(2), 212–221. https://doi.org/10.11594/jesi.02.02.09

Mahdiyan, A. (2023). *Perekonomian dunia diprediksi akan dihantam resesi tahun 2023, bagaimana dengan pembangunan infrastruktur?* https://kpbu.kemenkeu.go.id/read/1173-1508/umum/kajian-opini-publik/perekonomian-dunia-diprediksi-akan-dihantam-resesi-tahun-2023-bagaimana-dengan-pembangunan-infrastruktur

Qomariyah, S., Iriawan, N., & Fithriasari, K. (2019). Topic modeling Twitter data using Latent Dirichlet Allocation and Latent Semantic Analysis. *AIP Conference Proceedings*, *2194*(April 2017). https://doi.org/10.1063/1.5139825

Sutresno, S. A. (2023). Analisis Sentimen Masyarakat Indonesia Terhadap Dampak Penurunan Global Sebagai Akibat Resesi di Twitter. *Building of Informatics, Technology and Science (BITS)*, *4*(4), 1959–1966. https://doi.org/10.47065/bits.v4i4.3149

Trenquier, H. (2018). *Improving Semantic Quality of Topic Models for Forensic Investigations by*. 1–24.

Zakeri, B., Paulavets, K., Barreto-Gomez, L., Echeverri, L. G., Pachauri, S., Boza-Kiss, B., Zimm, C., Rogelj, J., Creutzig, F., Ürge-Vorsatz, D., Victor, D. G., Bazilian, M. D., Fritz, S., Gielen, D., McCollum, D. L., Srivastava, L., Hunt, J. D., & Pouya, S. (2022). Pandemic, War, and Global Energy Transitions. *Energies*, *15*(17), 1–23. https://doi.org/10.3390/en15176114