ANALISIS KATA KUNCI YANG RELEVAN DENGAN GAS RUMAH KACA DI INDONESIA MENGGUNAKAN *TEXT MINING* DAN TEORI LUHN

LAPORAN TUGAS AKHIR

Oleh:
Kiagus Muhammad Arsyad
105219002

FAKULTAS SAINS DAN ILMU KOMPUTER
PROGRAM STUDI ILMU KOMPUTER
UNIVERSITAS PERTAMINA
AGUSTUS 2023

ANALISIS KATA KUNCI YANG RELEVAN DENGAN GAS RUMAH KACA DI INDONESIA MENGGUNAKAN *TEXT MINING* DAN TEORI LUHN

LAPORAN TUGAS AKHIR

Oleh:
Kiagus Muhammad Arsyad
105219002

FAKULTAS SAINS DAN ILMU KOMPUTER
PROGRAM STUDI ILMU KOMPUTER
UNIVERSITAS PERTAMINA
AGUSTUS 2023


Universitas
Pertamina

LEMBAR PENGESAHAN

Judul Tugas Akhir	:	Analisis Kata Kunci yang Relevan dengan Gas Rumah Kaca di Indonesia Menggunakan *Text Mining* dan Teori Luhn
Nama Mahasiswa	:	Kiagus Muhammad Arsyad
Nomor Induk Mahasiswa	:	105219002
Program Studi	:	Ilmu Komputer
Fakultas	:	Sains dan Ilmu Komputer
Tanggal Lulus Sidang Tugas Akhir	:	2 Agustus 2023

Jakarta, 31 Juli 2023

MENGESAHKAN

Pembimbing I	Pembimbing II
Dr. Tasmi, S.Si., M.Si.	Dr. Ariana Yunita
NIP. 116109	NIP. 116015

MENGETAHUI,

Ketua Program Studi

Ade Irawan, Ph.D

NIP. 116130

LEMBAR PERNYATAAN

Dengan ini saya menyatakan bahwa Tugas Akhir berjudul Analisis Kata Kunci yang Relevan dengan Gas Rumah Kaca di Indonesia Menggunakan *Text Mining* dan Teori Luhn ini adalah benar-benar merupakan hasil karya saya sendiri dan tidak mengandung materi yang ditulis oleh orang lain kecuali telah dikutip sebagai referensi yang sumbernya telah dituliskan secara jelas sesuai dengan kaidah penulisan karya ilmiah.

Apabila dikemudian hari ditemukan adanya kecurangan dalam karya ini, saya bersedia menerima sanksi dari Universitas Pertamina sesuai dengan peraturan yang berlaku.

Demi pengembangan ilmu pengetahuan, saya menyetujui untuk memberikan kepada Universitas Pertamina hak bebas royalti noneksklusif (*non-exclusive royalty-free right*) atas Tugas Akhir ini beserta perangkat yang ada. Dengan hak bebas royalti noneksklusif ini Universitas Pertamina berhak menyimpan, mengalih media/format-kan, mengelola dalam bentuk pangkatan data (*database*), merawat, dan mempublikasikan Tugas Akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 31 Juli 2023

Yang membuat pernyataan,

– Materai –

Kiagus Muhammad Arsyad

ABSTRAK

Kiagus Muhammad Arsyad. **105219002**. Analisis Kata Kunci yang Relevan dengan Gas Rumah Kaca di Indonesia Menggunakan *Text Mining* dan Teori Luhn.

Penyebab Gas Rumah Kaca (GRK) menjadi isu global yang mendesak untuk ditangani, termasuk di Indonesia. Penelitian ini bertujuan untuk mengidentifikasi faktor penyebab GRK di Indonesia melalui analisis data abstrak dari studi literatur lima tahun terakhir. Pendekatan yang digunakan melibatkan teknik *text mining*, *clustering*, dan seleksi menggunakan teori Luhn. Dengan menggunakan algoritma K-Means dalam teknik *clustering*, data abstrak dikategorikan menjadi beberapa klaster berdasarkan kesamaan konten. Hasil analisis menunjukkan bahwa aspek yang memiliki relevansi GRK di Indonesia meliputi diantaranya konsumsi listrik, produksi energi, penggunaan energi, dan frasa yang lainnya. Metode seleksi menggunakan teori Luhn juga mampu mengidentifikasi kata-kata kunci yang paling relevan dan mewakili relevansi GRK di Indonesia. Pada penelitian ini juga memberikan pemahaman yang cukup tentang faktor-faktor yang berhubungan pada aspek GRK di Indonesia dan memberikan informasi penting bagi pembuat kebijakan, para ahli, para pembaca untuk mengembangkan strategi mitigasi yang efektif.

Kata kunci: faktor, gas rumah kaca, kata kunci, nlp, teori luhn, *text mining*

ABSTRACT

Kiagus Muhammad Arsyad. **105219002**. Analysis of Relevant Keywords Related to Greenhouse Gas in Indonesia Using Text Mining and Luhn's Theory.

The causes of Greenhouse Gas (GHG) have become a pressing global issue to be addressed, including in Indonesia. This study aims to identify the factors causing GHG in Indonesia through the analysis of abstract data from literature studies in the last five years. The approach used involves text mining, clustering, and selection techniques using Luhn's theory. By employing the K-Means algorithm in the clustering technique, the abstract data is categorized into several clusters based on content similarity. The results of the analysis show that aspects relevant to GHG in Indonesia include electricity consumption, energy production, energy use, and other phrases. The selection method using Luhn's theory also successfully identifies the most relevant keywords representing the relevance of GHG in Indonesia. This study provides a comprehensive understanding of the factors related to GHG aspects in Indonesia and offers important information for policymakers, experts, and readers to develop effective mitigation strategies.

Keywords: factors, greenhouse gas, keywords, luhn's theory, nlp, text mining

KATA PENGANTAR

Dengan memanjatkan segala puji syukur kehadirat Allah SWT yang telah melimpahkan rahmat, taufik, dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi berupa laporan Tugas Akhir (TA) yang berjudul **"Analisis Kata Kunci Faktor Penyebab Gas Rumah Kaca di Indonesia Melalui Pendekatan Kuantitatif dan Kualitatif pada Studi Literatur dengan *Text Mining* dan Teori Luhn"**, sebagai salah satu persyaratan dalam menyelesaikan masa belajar di Program Studi S1 Ilmu Komputer, Fakultas Sains dan Ilmu Komputer Universitas Pertamina. Penulis menyadari bahwa laporan TA ini tidak akan terselesaikan tanpa adanya dukungan, bimbingan, bantuan, dan nasihat dari berbagai pihak selama penyusunan laporan TA ini. Pada kesempatan ini, penulis ingin menyampaikan terima kasih setulus-tulusnya kepada:

1. Bapak Ade Irawan, Ph.D selaku Ketua Program Studi Ilmu Komputer Universitas Pertamina dan juga sebagai dosen penguji pertama yang telah memberikan masukan dan saran pada skripsi ini.
2. Bapak Rangga Ganzar Noegraha, Ph.D selaku dosen penguji kedua yang telah memberikan masukan dan saran pada skripsi ini.
3. Ibu Dr. Tasmi, S.Si, M.Si. dan Ibu Dr. Ariana Yunita selaku dosen pembimbing pertama dan dosen pembimbing kedua pada skripsi ini yang telah memberikan bimbingan, motivasi, dan arahan selama penyusunan laporan tugas akhir dari awal hingga akhir saat ini.
4. Seluruh staf pengajar yang telah memberikan ilmu pengetahuan yang tak ternilai selama penulis menempuh pendidikan di Universitas Pertamina.
5. Seluruh *civitas academica* di lingkungan Universitas Pertamina yang telah menunjang layanan dan dukungan selama proses pembelajaran di Universitas Pertamina.
6. Keluarga, saudara, dan teman-teman penulis yang telah memberikan doa, kasih sayang, motivasi, dan dorongan di setiap langkah perjuangan penulis.

Penulis menyadari bahwa laporan TA ini belum sempurna dan banyak kekurangan, oleh karena itu, segala saran dan kritik yang membangun diharapkan oleh penulis sebagai penyempurna penulisan laporan ini di masa yang mendatang. Semoga laporan TA ini dapat memberikan manfaat bagi penulis dan para pembaca.

Jakarta, 31 Juli 2023

Kiagus Muhammad Arsyad

DAFTAR ISI

KATA PENGANTAR . . . . . **iii**

DAFTAR ISI . . . . . **iv**

DAFTAR TABEL . . . . . **vi**

DAFTAR GAMBAR . . . . . **vii**

BAB I PENDAHULUAN **1**

1.1 Latar Belakang 1

1.2 Rumusan Masalah 2

1.3 Batasan Masalah 2

1.4 Tujuan penelitian 2

1.5 Manfaat penelitian 2

BAB II TINJAUAN PUSTAKA **4**

2.1 Gas Rumah Kaca (GRK) 4

2.2 *Text Mining* 5

2.2.1 *Natural Language Processing* (NLP) 5

2.2.2 *Term Frequency-Inverse Document Frequency* (TF-IDF) 6

2.2.3 *Principal Component Analysis* (PCA) 7

2.2.4 K-Means 7

2.3 Teknik Kitchenham 7

2.4 *Luhn's Theory* 8

BAB III METODE PENELITIAN **11**

3.1 Tahapan Penelitian 11

3.2 *Data Collection* 13

3.3 Tahapan *Text Mining* dan Interpretasinya 14

BAB IV HASIL DAN PEMBAHASAN **16**

4.1 *Preprocessing Data* 16
4.2 Hasil Pengelompokkan Kata Kunci berdasarkan teknik *clustering* 17
4.3 Hasil Pemilihan Kata Kunci berdasarkan Teori Luhn 21
4.4 Hasil Kata Kunci yang Didapatkan Berdasarkan Teknik *Clustering* dan Seleksi Teori Luhn 24

BAB V KESIMPULAN DAN SARAN **27**
5.1 Kesimpulan 27
5.2 Saran 27

DAFTAR PUSTAKA **29**

LAMPIRAN A Lampiran 1. Tren Total Emisi GRK Indonesia **33**

LAMPIRAN B Lampiran 2. Frasa dari Kata Kunci yang Didapatkan **34**

DAFTAR TABEL

Tabel 4.1 Hasil Kata Kunci yang Terpilih Berdasarkan Teknik *Clustering* dan Implementasi Teori Luhn 25

DAFTAR GAMBAR

Gambar 2.1 Diagram *Word Frequency* 9

Gambar 3.1 Diagram Alir Penelitian 11

Gambar 3.2 Tahapan *Data Collection* hingga hasil interpretasi 12

Gambar 3.3 Tahapan proses pemilihan data studi literatur 13

Gambar 3.4 Alur tahapan saat dilakukan *text mining* hingga hasil interpretasi 14

Gambar 4.1 Proses mendapatkan kata-kata dasar dari data studi literatur 16

Gambar 4.2 Visualisasi hasil dari PCA 18

Gambar 4.3 *Elbow Plot* untuk menenentukan nilai klaster dengan K-Means 18

Gambar 4.4 Metrik Evaluasi nilai K dengan *Davies-Bouldin index* dan *library ElbowVisualizer* 19

Gambar 4.5 Klasterisasi Pewarnaan dengan K-Means 20

Gambar 4.6 *Silhouette Plot* untuk K-Means *Clustering* 20

Gambar 4.7 Representasi visual frekuensi data dengan *word cloud* 22

Gambar 4.8 Dua puluh kata dasar dengan frekuensi terbesar 22

Gambar 4.9 *Preview* kata-kata dengan frekuensi terbesar 23

Gambar 4.10 *Preview* kata-kata dengan frekuensi terendah 23

Gambar 4.11 Hasil Kata Kunci Berdasarkan Hasil Analisis Manual dan Implementasi Teori Luhn 24

Gambar 1.1 Total emisi GRK di Indonesia 2010-2019 33

Gambar 2.1 Hasil frasa yang dapat dibentuk (1) 34

Gambar 2.2 Hasil frasa yang dapat dibentuk (2) 35


Universitas
Pertamina

BAB I
PENDAHULUAN

1.1 Latar Belakang

Emisi gas rumah kaca (GRK) di Indonesia diperkirakan meningkat pada periode 2021-2030. Informasi tersebut diperoleh dari artikel DataIndonesia.Id yang ditulis oleh Rizaty [1], pada artikel tersebut juga disebutkan bahwa emisi GRK nasional sudah mencapai 259,1 juta ton CO2 pada tahun 2021. Proyeksi emisi GRK tahun 2030 diprediksi akan meningkat sebesar 29,13% menjadi 334,6 juta ton CO2.

Selain itu, Indonesia akan terus meningkat diikuti dengan kemajuan teknologi, hal ini menyebabkan peningkatan kebutuhan energi [2]. Pertumbuhan penduduk ini berdampak pada penggunaan bahan bakar fosil, seperti pembakaran kendaraan bermotor dan kegiatan industri, yang menjadi penyumbang salah satu faktor emisi GRK [3]. Dampak dari peningkatan emisi GRK dan konsumsi energi di dunia juga sangat signifikan terhadap lingkungan, seperti kenaikan suhu global, perubahan iklim ekstrem, serta perubahan pola cuaca [4]. Hal ini perlu diwaspadai mengingat dampak dari emisi GRK yang cukup membahayakan.

Di Indonesia, terdapat program yang dinamakan Indonesia's FOLU Net Sink 2030, yang merupakan sebuah inisiatif dari pemerintah Indonesia untuk mencapai keseimbangan antara emisi dan penyerapan karbon pada tahun 2030. Dalam rangka mengatasi isu perubahan iklim, Indonesia berkomitmen dalam memperkuat sektor kehutanan, pertanian, dan penggunaan lahan lainnya sebagai upaya mencapai "*net sink*" yaitu penyerapan karbon yang lebih banyak [5].

Inovasi, upaya, dan ide yang dihasilkan dari penelitian yang membahas tentang emisi gas rumah kaca (GRK) di Indonesia juga diperlukan saat ini dalam memberikan kontribusi yang lebih efektif dalam mencapai target Indonesia's FOLU Net Sink 2030. Dalam rangka mencari solusi terhadap masalah emisi GRK di Indonesia, pemanfaatan teknologi dan kecerdasan buatan, seperti *Natural Language Processing* (NLP) dan *text mining*, dapat membantu mengatasi tantangan tersebut [6, 7]. Teks-teks dari berbagai sumber, seperti jurnal penelitian, artikel, dan laporan yang membahas mengenai GRK dan terdapat hubungannya di Indonesia dapat dijadikan sebagai data. Data teks tersebut dapat diolah dan dianalisis dengan efisien menggunakan proses *text mining* dan NLP dalam mengidentifikasi faktor-faktor penyebab GRK yang relevan. Selain itu, terdapat pendekatan menggunakan teori Luhn [8] sebagai *text summarizer* [9] atau dalam pemilihan kata kunci sesuai dengan konsep yang ingin dicari [10], sehingga dapat meningkatkan kualitas dan validitas hasil dari proses tersebut.

Oleh sebab itu, penulis ingin menerapkan pendekatan gabungan berupa kuantitatif dan kualitatif yang dapat menentukan kata-kata kunci yang relevan mengenai GRK di Indonesia berdasarkan data yang tersedia dengan didukung pencarian studi literatur. Diharapkan dari penelitian ini dapat memberikan pengetahuan baru dan ide terkait solusi yang dapat dilakukan untuk mengurangi emisi GRK. Selain itu juga dapat mendukung upaya kepada pembaca, para pakar, peneliti, instansi maupun peme-

rintah dalam mengoptimalkan kebijakan dan strategi untuk mengatasi perubahan iklim secara lebih efisien dan efektif khususnya dalam aspek GRK di Indonesia.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, rumusan masalah pada penelitian ini yaitu bagaimana cara menentukan kata-kata kunci yang memiliki relevansi faktor penyebab gas rumah kaca di Indonesia berdasarkan data studi literatur dengan *text mining* berupa teknik *clustering* dan teori Luhn?

1.3 Batasan Masalah

Adapun batasan masalah pada penelitian ini ialah menggunakan data berupa studi literatur yang merupakan artikel atau penelitian lima tahun terakhir sejak tahun 2018 yang didapatkan melalui *search engine* Scopus yang mengulas terkait GRK di negara Indonesia.

1.4 Tujuan penelitian

Tujuan dari penelitian ini yaitu untuk menentukan kata-kata kunci yang memiliki relevansi faktor penyebab gas rumah kaca di Indonesia berdasarkan data studi literatur dengan *text mining* berupa teknik *clustering* dan teori Luhn.

1.5 Manfaat penelitian

Penelitian ini diharapkan dapat memberikan manfaat yang signifikan dalam menentukan faktor penyebab emisi Gas Rumah Kaca (GRK) di Indonesia berdasarkan keterbaruan dari salah satu basis data literatur ilmiah yaitu Scopus. Dengan menerapkan pendekatan gabungan yaitu kuantitatif dan kualitatif menggunakan *text mining* dengan NLP serta K-Means dan teori Luhn, penelitian ini diharapkan juga dapat memberikan kemudahan dan efisiensi dalam mengidentifikasi kata kunci yang relevan mengenai GRK secara valid berdasarkan fakta dan data yang terverifikasi. Selain itu, hasil penelitian ini dapat menjadi dasar untuk mengembangkan kebijakan dan strategi yang lebih baik berdasarkan aspek relevansi dalam menghadapi tantangan perubahan iklim dan menjaga keberlanjutan lingkungan bagi masa depan yang lebih baik terutama dalam kasus GRK di Indonesia.


Universitas
Pertamina

BAB II
TINJAUAN PUSTAKA

2.1 Gas Rumah Kaca (GRK)

Perubahan iklim telah menjadi isu keamanan internasional yang penting dan nontradisional. Emisi gas rumah kaca (GRK) yang berlebihan merupakan salah satu penyebab utama dari meningkatnya permasalahan iklim. Perubahan iklim tidak hanya secara langsung mempengaruhi kesehatan populasi dengan meningkatkan frekuensi dan intensitas gelombang panas, kekeringan, dan curah hujan yang tinggi, tetapi juga secara tidak langsung dengan meningkatkan polusi udara, mempercepat penyebaran vektor penyakit, serta mempengaruhi keamanan pangan dan kesehatan mental. Emisi GRK yang berlebihan menjadi faktor yang memperburuk perubahan iklim dan dampaknya terhadap berbagai aspek kehidupan manusia [11].

Gas rumah kaca merujuk pada gas-gas yang hadir di atmosfer, baik secara alami maupun sebagai hasil aktivitas manusia (antropogenik), yang mampu menyerap dan memancarkan kembali radiasi inframerah [12]. Ketika permukaan bumi menerima radiasi matahari dalam bentuk gelombang pendek, sebagian besar radiasi ini dipancarkan kembali ke atmosfer sebagai radiasi gelombang panjang (infra merah). Gas rumah kaca yang terdapat di lapisan atmosfer yang dekat dengan permukaan bumi menyerap radiasi gelombang panjang ini, menyebabkan peningkatan suhu yang tinggi yang dikenal sebagai efek rumah kaca. Peningkatan suhu ini terjadi akibat perubahan kondisi dan komposisi atmosfer yang mengelilingi planet ini [13].

Dalam era saat ini, masalah lingkungan telah menjadi pembahasan utama baik di negara-negara yang sedang berkembang maupun negara-negara maju karena adanya kerusakan lingkungan. Hal ini juga menimbulkan pertanyaan tentang pemanasan global dan perubahan iklim, yang terutama disebabkan oleh emisi gas rumah kaca, kadang-kadang terkait dengan penyebab alami seperti pergeseran benua, aktivitas gunung berapi, radiasi matahari, dan arus laut, serta aktivitas manusia langsung maupun tidak langsung yang mempengaruhi komposisi atmosfer global dan variasi lingkungan [4]. Para peneliti telah berargumen bahwa peningkatan aktivitas manusia akibat perkembangan industrialisasi, pertumbuhan populasi global, dan kebutuhan untuk mengatasi perubahan tersebut adalah penyebab utama perubahan iklim. Selain itu, aktivitas manusia seperti deforestasi pertanian dan komersial, pembakaran bahan bakar fosil, serta perubahan penggunaan lahan akibat pertumbuhan populasi juga memberikan kontribusi yang signifikan terhadap peningkatan emisi gas rumah kaca [14].

Menurut Khairunnisa Musari & Sayah [15], dalam rangka penyelesaian masalah gas rumah kaca (GRK), beberapa hal yang perlu diperhatikan adalah upaya melawan perubahan iklim, prioritas nasional, transformasi kebijakan, menciptakan lingkungan yang mendukung, dan investasi keuangan yang diperlukan, yang semuanya harus menjadi bagian dari agenda nasional. Selain itu dengan perkembangan era informasi saat ini, salah satu upaya dengan pemanfaatan teknologi atau kecerdasan buatan (*Artificial Intelligence*/AI) menjadi hal yang penting karena dapat digunakan untuk pemantauan, analisis, dan pengelolaan data terkait emisi GRK. Sementara itu, dapat membantu dalam

mengoptimalkan kebijakan dan strategi untuk mengurangi emisi GRK secara efisien. Dengan memanfaatkannya secara holistik, diharapkan dapat memberikan penyelesaian masalah GRK dan perubahan iklim dapat tercapai dengan lebih efektif dan efisien.

2.2 *Text Mining*

Studi literatur mengenai *text mining* telah menjadi area penelitian yang semakin populer dalam ilmu data dan pengolahan bahasa alami. *Text mining* merupakan teknik yang digunakan untuk mengekstraksi informasi berharga dari data teks yang tidak terstruktur, seperti artikel jurnal, berita, dan dokumen teks lainnya. Teknik ini mencakup beberapa tahapan, termasuk pengumpulan data teks, pemrosesan teks untuk menghapus karakter-karakter tidak penting, dan analisis data untuk mengidentifikasi pola, topik, atau informasi penting dari teks tersebut [7].

Dalam penerapannya, *text mining* telah digunakan dalam berbagai bidang. Misalnya, dalam analisis sentimen, *text mining* dapat digunakan untuk mengekstraksi sentimen atau perasaan dari teks yang ditulis oleh pengguna dalam media sosial atau ulasan produk [16]. Selain itu, dalam pengelompokan topik, *text mining* dapat membantu mengelompokkan dokumen-dokumen teks berdasarkan tema atau topik yang sama.

Dalam penelitian ini, *text mining* akan digunakan sebagai salah satu teknik dalam *preprocessing dataset* berupa teks untuk mendapatkan kata kunci yang relevan dengan permasalahan gas rumah kaca (GRK) di Indonesia. Teknik *text mining* akan membantu dalam ekstraksi fitur teks dari data teks yang tidak terstruktur, yakni judul dan abstrak jurnal penelitian yang terdapat pada *dataset*. Proses *text mining* akan melibatkan langkah-langkah seperti dengan teknik *stopwords, lemmatization, punctuation, word frequency* dari ekstraksi *unigram* yang akan menjadi kata-kata dasar dari *dataset* tersebut.

2.2.1 *Natural Language Processing* (NLP)

Bahasa Pemrosesan Alami atau yang biasa disebut *Natural Language Processing* (NLP) adalah bidang keilmuan yang berfokus pada pemahaman dan penerapan bahasa manusia dalam bentuk teks atau ucapan oleh komputer. NLP telah mengalami perkembangan pesat dalam beberapa dekade terakhir, berkat kemajuan teknologi komputasi dan kemampuan mesin dalam memproses data secara lebih efektif. Salah satu tantangan utama dalam NLP adalah mengenali dan memahami struktur dan makna dari teks bahasa manusia, karena bahasa manusia seringkali ambigu dan penuh dengan variasi. Beberapa aplikasi praktis dari NLP termasuk mesin penerjemah, asisten virtual, analisis sentimen, pengenalan suara, dan masih banyak lagi. NLP telah menjadi bidang penelitian yang sangat menarik dan relevan dalam era informasi digital yang sedang berkembang pesat [6].

Salah satu pendekatan populer dalam NLP adalah menggunakan teknik pembelajaran mesin (machine learning) untuk mengajarkan komputer bagaimana memahami dan memproses bahasa manusia. Teknik-teknik ini melibatkan penggunaan model statistik dan algoritma yang kompleks untuk mengidentifikasi pola, keterkaitan, dan makna dalam teks bahasa manusia. Contoh teknik pembelajaran mesin yang sering digunakan dalam NLP termasuk model berbasis aturan, model berbasis vektor, dan model berbasis jaringan saraf tiruan [17]. Dalam beberapa tahun terakhir, penggunaan jaringan

saraf tiruan, khususnya dalam bentuk transformasi bahasa seperti BERT dan GPT-3, telah mencatat kemajuan signifikan dalam banyak tugas NLP, seperti penerjemahan mesin dan pemahaman bahasa alami [18].

Dalam penelitian ini, teknik pemrosesan bahasa alami (Natural Language Processing/NLP) akan digunakan dalam tahapan *text mining* pada data berupa teks untuk mendapatkan kata kunci yang relevan. Proses pada tahapan tersebut meliputi *preprocessing* dengan menerapkan langkah-langkah seperti menghapus kata umum dan tanda baca, lematisasi, serta mendapatkan kata-kata yang relevan dari teks. Selain itu, teknik ekstraksi fitur seperti metode Term Frequency-Inverse Document Frequency (TF-IDF) digunakan untuk mengidentifikasi kata-kata yang paling penting dan mewakili konten teks secara numerik [19]. Hasil dari penggunaan NLP dalam pemrosesan data teks ini akan digunakan sebelum menentukan hasil interpretasi yang diharapkan pada penelitian ini.

2.2.2 *Term Frequency-Inverse Document Frequency* (TF-IDF)

TF-IDF (*Term Frequency-Inverse Document Frequency*) merupakan salah satu teknik yang sering digunakan dalam pengolahan teks dan *Information Retrieval* (IR). Teknik ini berfungsi untuk memberikan bobot pada setiap kata dalam dokumen berdasarkan frekuensi kemunculan kata tersebut dalam dokumen itu sendiri (TF) dan kebalikannya, yakni berdasarkan keberbedaan jumlah dokumen yang mengandung kata tersebut di dalam seluruh koleksi dokumen (IDF) [20]. Dengan menggunakan TF-IDF, kata-kata yang sering muncul dalam suatu dokumen tetapi jarang muncul di dokumen lain akan memiliki bobot yang tinggi, sehingga dapat dianggap sebagai kata kunci atau kata yang lebih penting dalam konteks dokumen tersebut. Penggunaan TF-IDF telah terbukti efektif dalam berbagai aplikasi seperti klasifikasi teks, analisis sentimen, dan sistem rekomendasi, serta telah menjadi salah satu teknik yang populer dalam mengolah teks untuk mendapatkan informasi yang bermakna [21].

Adapun TF-IDF *Vectorizer* adalah salah satu fungsi yang digunakan dalam pengolahan teks untuk mengimplementasikan metode TF-IDF. TF-IDF *Vectorizer* berfungsi untuk mengubah teks menjadi vektor numerik berdasarkan nilai TF-IDF dari setiap kata dalam teks [21]. Proses ini memungkinkan data teks yang kompleks dan tidak terstruktur diubah menjadi representasi numerik yang dapat diolah lebih lanjut dengan algoritma pembelajaran mesin atau analisis statistik.

Penggunaan TF-IDF *Vectorizer* sangat berguna dalam analisis teks dan NLP, terutama dalam pemrosesan data besar yang mengandung banyak dokumen atau teks. Dengan menggunakan TF-IDF *Vectorizer*, dapat dilakukan suatu pengelompokan teks, analisis sentimen, deteksi topik, dan sistem rekomendasi secara efisien. Selain itu, dengan menerapkan TF-IDF *Vectorizer* dalam penelitian ini mengenai emisi gas rumah kaca (GRK) di Indonesia, data teks abstrak akan dikonversi menjadi representasi vektor TF-IDF dengan memperhitungkan bobot kata dalam setiap abstrak. Kata-kata kunci yang relevan dalam data teks yang berkaitan dengan GRK dan menganalisis tingkat signifikansinya dalam dokumen tersebut.

2.2.3 *Principal Component Analysis* (PCA)

PCA (*Principal Component Analysis*) adalah sebuah metode statistik yang digunakan untuk mengurangi dimensi dari data dengan mempertahankan informasi yang paling penting. PCA digunakan untuk mengidentifikasi pola dalam data, dan mengubah data tersebut menjadi bentuk yang lebih mudah dipahami dan diinterpretasikan [22].

PCA lebih sering digunakan dalam analisis data numerik, namun PCA juga dapat digunakan dalam analisis teks. Dalam analisis teks, PCA digunakan untuk mengurangi dimensi dari data teks dengan mempertahankan informasi yang paling penting [23]. PCA dapat digunakan untuk mengidentifikasi pola dalam data teks, seperti pola kata yang sering muncul bersama-sama atau pola topik yang terkait. Dengan mengurangi dimensi dari data teks, PCA dapat mempermudah analisis dan interpretasi data teks.

2.2.4 K-Means

K-Means adalah salah satu algoritma *clustering* yang paling populer digunakan dalam analisis data. Algoritma ini digunakan untuk membagi data menjadi beberapa kelompok atau klaster berdasarkan kesamaan fitur atau karakteristik tertentu [24]. Dalam K-Means, data dikelompokkan berdasarkan jarak antara titik data dan *centroid*, yang merupakan pusat dari setiap klaster. Tujuan dari algoritma ini adalah untuk meminimalkan varians dalam setiap klaster dan memaksimalkan jarak antara klaster.

Salah satu penerapan K-Means yang sering digunakan adalah dalam analisis teks. Dalam analisis teks, K-Means digunakan untuk mengelompokkan dokumen berdasarkan kesamaan topik atau kata kunci tertentu [25]. Misalnya, jika kita memiliki kumpulan dokumen tentang olahraga, K-Means dapat digunakan untuk mengelompokkan dokumen berdasarkan jenis olahraga atau topik tertentu seperti sepak bola, basket, atau tenis. Dalam hal ini, K-Means dapat membantu mengidentifikasi pola dan tren dalam data teks yang besar dan kompleks.

K-Means dapat digunakan untuk memproses hasil dari analisis PCA (Principal Component Analysis). PCA adalah teknik statistik yang digunakan untuk mengurangi dimensi data dengan mempertahankan informasi yang paling penting. Setelah PCA diterapkan pada data, K-Means dapat digunakan untuk mengelompokkan data menjadi beberapa klaster berdasarkan kesamaan fitur atau karakteristik tertentu [25]. Hasil dari analisis K-Means dapat membantu mengidentifikasi pola dan tren dalam data yang besar dan kompleks, serta memudahkan interpretasi hasil analisis.

2.3 Teknik Kitchenham

Teknik Kitchenham merujuk pada metode sistematis literature review (SLR) yang dikembangkan oleh Barbara Kitchenham, seorang peneliti terkemuka di bidang rekayasa perangkat lunak [26]. Pedoman Kitchenham untuk melakukan SLR telah menjadi sangat diakui dan sering disebut sebagai ”teknik Kitchenham”.

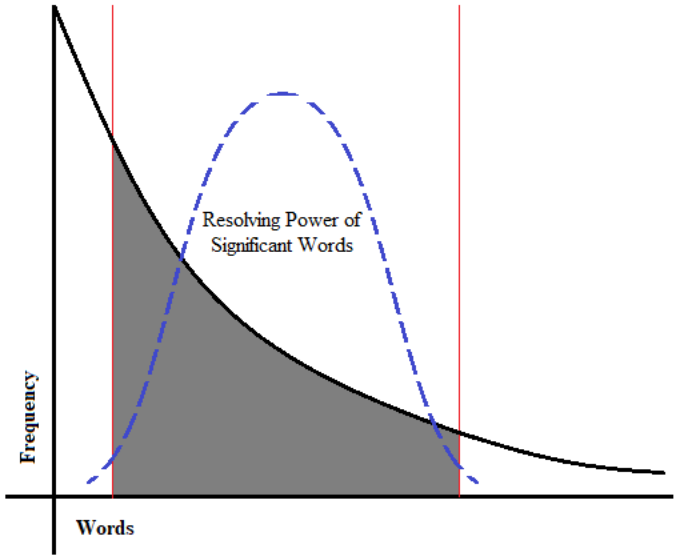
Teknik Kitchenham melibatkan proses dalam mengidentifikasi, mengevaluasi, dan mensintesis semua studi yang relevan pada pertanyaan atau topik penelitian tertentu. Proses ini meliputi pencarian yang komprehensif dan sistematis dari beberapa database dan sumber lainnya, pengembangan kriteria inklusi dan eksklusi, dan teknik analisis data untuk memastikan bahwa hasilnya dapat diandalkan dan tidak bias. Pedoman Kitchenham telah banyak diadopsi dalam penelitian rekayasa perangkat lunak dan telah berkontribusi pada pengembangan rekayasa perangkat lunak berbasis bukti (EBSE) [26, 27]. Teknik Kitchenham telah menjadi metode standar untuk melakukan SLR dalam rekayasa perangkat lunak dan telah membantu meningkatkan kualitas dan keandalan penelitian di bidang tersebut.

Dalam konteks penggunaan teknik Kitchenham, terdapat penggunaan operator logika seperti *AND*, *OR*, dan *NOT* dalam pencarian studi yang relevan. Operator logika tersebut digunakan untuk menggabungkan kata kunci pencarian dan memperluas atau mempersempit cakupan pencarian. Misalnya, jika ingin mencari studi yang relevan dengan topik "*greenhouse gas*" dapat menggunakan operator logika AND untuk menggabungkan kata kunci "*greenhouse*" dan "*gas*". Dengan demikian, pencarian akan memperlihatkan studi yang hanya relevan dengan kedua kata kunci tersebut.

Selain itu juga, penggunaan asterisk (*) dalam pencarian studi yang relevan juga dapat mencakup variasi kata yang terkait dengan akar kata tertentu. Misalnya, jika menggunakan kata kunci "*communicat**", pencarian akan mencakup studi yang menggunakan kata "*communicate*", "*communication*", "*communicating*", dan sebagainya. Dengan menggunakan asterisk, akan dapat memperluas cakupan pencarian dan memastikan bahwa semua studi yang relevan ditemukan, terutama jika ada variasi dalam cara menyebutkan kata kunci tertentu.

2.4 *Luhn's Theory*

Teori yang digagas oleh H. P. Luhn, dalam penelitiannya yaitu penciptaan otomatis abstrak literatur menggunakan mesin pemrosesan data [8]. Implementasi Luhn melibatkan analisis teks lengkap dari sebuah artikel dan pemilihan kalimat dan frasa kunci untuk membuat ringkasan singkat dari konten artikel. Mesin pemrosesan data IBM 704 digunakan untuk menghitung ukuran relatif signifikansi untuk kata dan kalimat individu, yang kemudian digunakan untuk memilih informasi paling penting untuk disertakan dalam abstrak. Implementasi ini terbukti efektif dalam menciptakan abstrak yang melayani tujuan abstrak konvensional dan dapat menghemat waktu dan usaha bagi pembaca literatur teknis.



Gambar 2.1. Diagram *Word Frequency*

Hal yang menjadi sebagai kata kunci dalam penerapan metode otomatis untuk membuat abstrak literatur menggunakan mesin pemrosesan data adalah kata-kata yang paling relevan dan penting dalam artikel yang sedang dianalisis. Kata-kata ini biasanya dipilih berdasarkan topik atau subjek dari artikel tersebut. Dalam *common words* seperti *pronouns*, *prepositions*, dan *articles* dihapus dari daftar kata-kata yang dianalisis karena kata-kata ini tidak memberikan informasi yang signifikan tentang topik atau subjek dari artikel. Sebaliknya, kata-kata yang lebih spesifik dan relevan seperti kata benda dan kata kerja digunakan sebagai kata kunci untuk memilih kalimat-kalimat kunci dalam artikel yang akan digunakan untuk membuat abstrak.

Terdapat penelitian yang dilakukan oleh Rahmah et al [10] dalam menerapkan teori Luhn, dengan Hukum Zipf yang menyatakan bahwa frekuensi kemunculan sebuah kata dalam teks berbanding terbalik dengan peringkatnya dalam daftar frekuensi kata. Dalam konteks pembuatan abstrak otomatis, hukum Zipf dapat digunakan untuk memilih kata-kata kunci yang paling relevan dan penting dalam artikel yang sedang dianalisis. Selain itu, juga membahas tentang skema Bradford [8], yang merupakan metode untuk mengelompokkan jurnal-jurnal ilmiah berdasarkan topik atau subjek yang sama, dengan adanya *lower cut* dan *upper cut* akan dipilih kata-kata yang diseleksi sebagai tema atau konsep yang dicari sebagaimana yang ditunjukkan pada Gambar 2.1 [8]. Metode tersebut digunakan untuk membantu pembaca literatur teknis menemukan jurnal-jurnal yang relevan dengan topik atau subjek yang sedang mereka teliti [10].

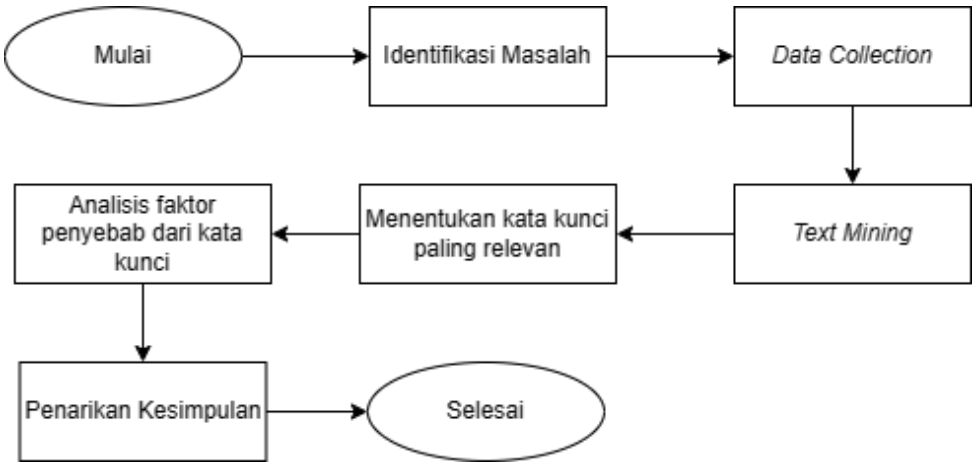

Universitas
Pertamina

BAB III
METODE PENELITIAN

3.1 Tahapan Penelitian

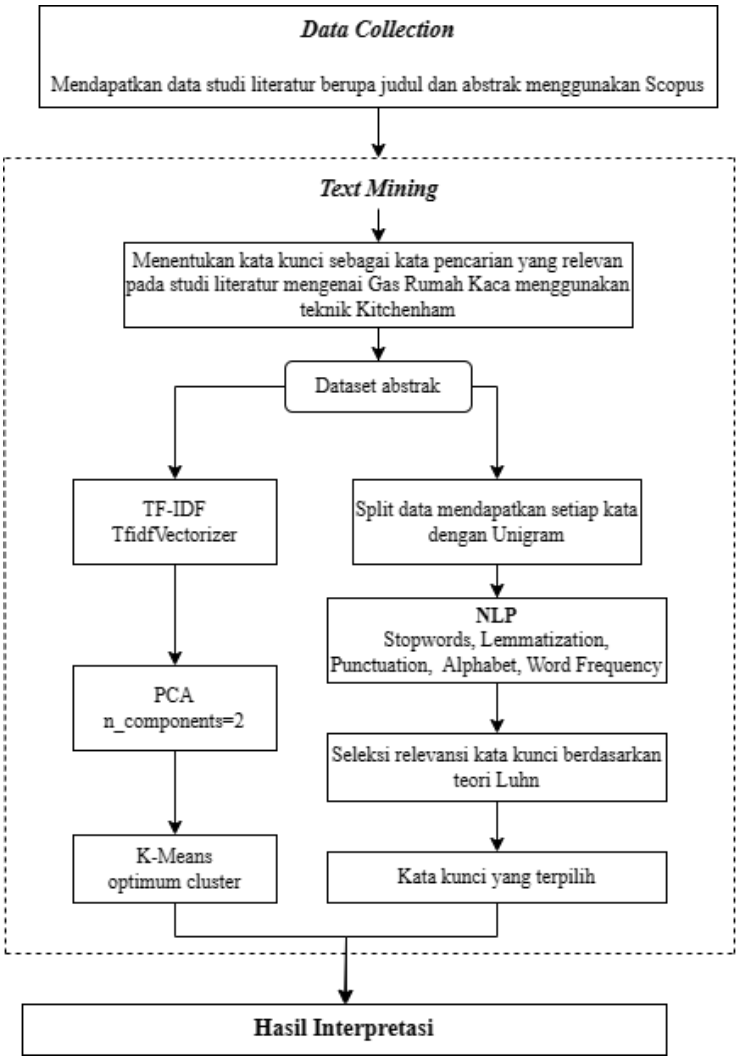
Penelitian ini mengikuti diagram alir yang terdiri dari beberapa tahapan. Tahap pertama adalah mengidentifikasi masalah terkait permasalahan Gas Rumah Kaca (GRK) di Indonesia. Selanjutnya, dilakukan tahap *data collection* dengan mengumpulkan data abstrak dari *database* jurnal penelitian dengan platform Scopus. Setelah itu, data tersebut akan diproses menggunakan teknik *text mining* untuk mendapatkan kata-kata kunci yang relevan.

Selanjutnya, berdasarkan kata-kata dasar yang dihasilkan dari *text mining*, dilakukan pemilihan kata kunci dari kata-kata dasar yang didapatkan dengan mencari relevansi dengan faktor-faktor penyebab GRK, serta hasil analisis menggunakan metode yang dipakai yaitu TF-IDF, PCA, K-Means, dan teori Luhn. Tahap akhir adalah menarik kesimpulan dari hasil penelitian. Dengan demikian, diagram alir penelitian ini menggambarkan langkah-langkah yang sistematis untuk memahami faktor penyebab permasalahan GRK di Indonesia berdasarkan analisis kata kunci.



Gambar 3.1. Diagram Alir Penelitian

Pada tahapan ketika melakukan *data collection* hingga ke hasil interpretasi, dengan menerapkan skema yang pernah digunakan oleh Lai [28] dengan skemanya yaitu ”*social media to-concepts*”, secara rinci akan dijelaskan di alur tahapannya sebagai berikut:



Gambar 3.2. Tahapan *Data Collection* hingga hasil interpretasi

Berdasarkan tahapan yang digambarkan pada Gambar 3.2, hal yang pertama dilakukan yaitu identifikasi masalah terkait emisi gas rumah kaca (GRK) di Indonesia yang relevan untuk pencarian data yang dilanjutkan dengan studi literatur pada *dataset* yang mendukung untuk mengidentifikasi faktor-faktor penyebab GRK. Adapun pada pemilihan data akan diambil berdasarkan abstrak yang berhubungan dengan GRK di Indonesia, kemudian dilakukan analisis lebih detil terkait relevansi, topik spesifik, dan keterbaruan untuk menentukan teknik apa yang akan digunakan dalam pencarian kata-kata kunci dari *dataset* tersebut.

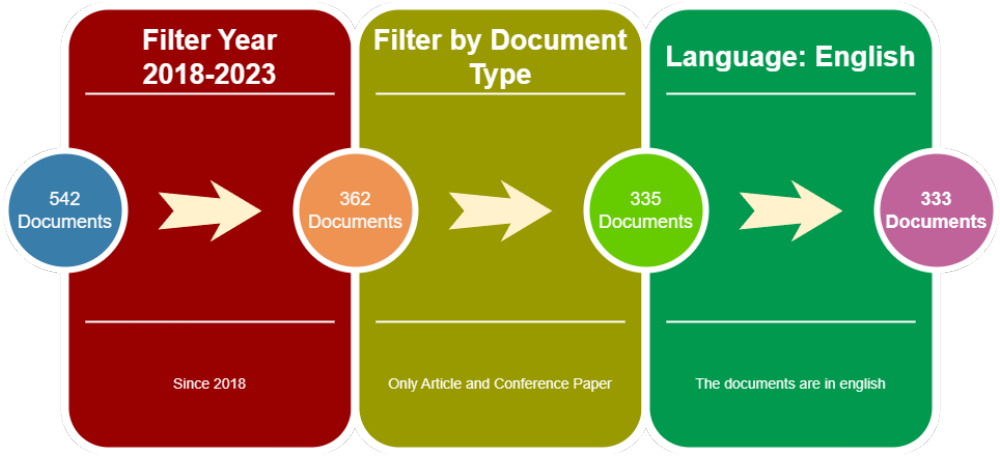
Selanjutnya, data yang telah dikumpulkan akan diproses menggunakan NLP dengan mengimplementasikan *splitting data* secara Unigram (N-grams $= 1$) untuk mendapatkan kata kunci per huruf dari *dataset*. Kata-kata kunci ini nantinya akan digunakan dalam analisis kualitatif. Dari banyaknya kata kunci tersebut yang memiliki frekuensi paling besar, akan dipilih kata-kata yang paling spesifik berdasarkan teori Luhn. Setelah semua tahapan di atas dilakukan, proses interpretasi akan dilakukan untuk mencari faktor-faktor penyebab GRK berdasarkan hasil yang telah diperoleh. Selanjutnya, akan ditemukan hasil upaya atau solusi dari faktor-faktor penyebab GRK yang telah dipilih berdasarkan kata kunci yang telah didapatkan.

3.2 *Data Collection*

Pada tahapan *data collection*, data yang diambil berdasarkan studi literatur yang telah didapatkan melalui *search engine* Scopus yang mengulas terkait GRK. Data yang diambil tersebut merupakan kumpulan abstrak di artikel atau penelitian yang dipilih berdasarkan pemilihan kata kunci dengan teknik Kitchenham [26]. Adapun dalam mendapatkan data studi literatur, diawali dengan mencari faktor-faktor penyebab gas rumah kaca terlebih dahulu dengan mendapatkan sumber referensi yang telah pernah membahas faktor penyebab gas rumah kaca. Adapun cara pencariannya terinspirasi dari penelitian yang dilakukan oleh [29] dengan melakukan teknik Kitchenham [26].

Pada pencarian studi literatur berupa jurnal atau artikel yang akan dipakai sebagai data penelitian ini, dilakukan pencarian di platform Scopus dengan implementasi Kitchenham yakni mendefinisikan objektif penelitian dengan kata kunci agar didapatkan sekumpulan dan abstrak yang relevan secara efektif. Dimulai dari aspek yang berkaitan dengan penyebab, maka kata kunci yang dipakai yaitu "*factor, feature, variable, cause, character, impact*", kemudian dicari berdasarkan aspek hasil yang ditimbulkan yaitu dengan kata kunci "*affect, contribute, product, generate, conduct*". Setelah itu, dari segi nama subjek yang dipakai mengenai GRK yakni "GHG, greenhouse gas", dan yang berdasarkan negara atau daerah spesifik yaitu di Indonesia. Maka, rangkaian kata kunci yang akan dipakai untuk ditulis di *search engine* Scopus ialah:

(factor OR featur* OR variabl* OR caus* OR character* OR impact) AND (affect* OR contribut* OR produc* OR generat* OR conduc*) AND (ghg OR greenhouse AND gas*) AND (indonesia OR indonesian)*



Gambar 3.3. Tahapan proses pemilihan data studi literatur

Setelah mendefinisikan kata kunci pada pencarian berdasarkan objektif penelitian, terdapat tahapan yang diperlukan dalam pemilihan studi literatur yang cocok sebagaimana yang telah digambarkan pada Gambar 3.3. Pada pencarian pertama, didapatkan sebanyak 542 dokumen yang tersedia. Agar data studi literatur yang dipakai merupakan dokumen yang baru, maka diambil berdasarkan terbitan publikasi 5 tahun saat ini yaitu tahun 2018 hingga saat ini, maka dihasilkan sebanyak 362 dokumen. Setelah itu, untuk mendapatkan jenis studi literatur yang spesifik, maka dipilih jenisnya yaitu *article* dan *conference paper* dan didapatlah sebanyak 335 dokumen. Pada tahap akhir, dari dokumen studi

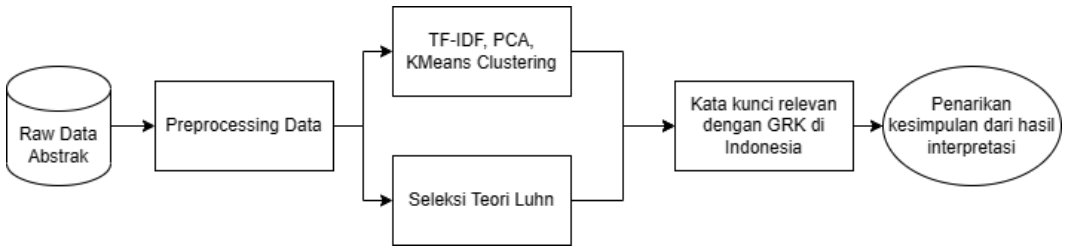
literatur yang didapat perlu dipilih hanya yang berbahasa Inggris saja, untuk pemahaman dan saat pemrosesan data nanti.

Berikut ini *query search* yang dihasilkan setelah semua pemilihan dokumen studi literatur tersebut dilakukan di platform Scopus:

TITLE-ABS-KEY (factor OR featur* OR variabl* OR caus* OR character* OR impact AND affect* OR contribut* OR produc* OR generat* OR conduc* AND ghg OR greenhouse AND gas* AND indonesia OR indonesian) AND PUBYEAR > 2017 AND PUBYEAR < 2024 AND (LIMIT-TO (LANGUAGE , "English")) AND (LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "cp"))*

Maka, setelah pendefinisian kata kunci dengan teknik Kitchenham, dilanjutkan dengan penyaringan dokumen di platform Scopus, sebanyak 333 dokumen, akan dipakai teks abstrak sebagai *dataset* dalam penelitian ini.

3.3 Tahapan *Text Mining* dan Interpretasinya



Gambar 3.4. Alur tahapan saat dilakukan *text mining* hingga hasil interpretasi

Pada tahapan ini, akan lebih banyak implementasi menggunakan NLP, menggunakan data yang berbentuk teks sekumpulan kalimat abstrak yang telah didapatkan, nantinya data tersebut dilakukan *preprocessing data* untuk menyortir teks agar dapat diproses dan siap untuk digunakan dalam analisis lebih lanjut secara baik. Setelah dilakukannya *preprocessing data*, akan dilakukan analisis 2 arah yaitu dengan teknik *clustering* dan teknik mencari kata kunci yang relevan.

Pada analisis *clustering*, proses tersebut melakukan representasi teks menggunakan TF-IDF, dilanjutkan dengan reduksi dimensi menggunakan PCA, kemudian melakukan K-Means clustering untuk mengelompokkan abstrak berdasarkan fitur-fitur yang dihasilkan dari PCA. Hasil akhirnya adalah pengelompokan abstrak ke dalam beberapa klaster berdasarkan kesamaan fitur-fitur TF-IDF yang telah direduksi menggunakan PCA.

Kemudian pada analisis seleksi kata-kata dasar yang menjadi kata kunci yang memiliki relevansi dengan faktor penyebab GRK, akan diseleksi kata kunci apa yang sesuai berdasarkan cara dari teori Luhn. Nantinya dari data berupa kata kunci yang terpilih atau terseleksi tersebut, serta hasil dari *clustering* sebelumnya, akan dipakai di dalam mencari hasil interpretasi berupa kata kunci apa yang sering dibahas sebagai penyebab GRK di Indonesia, dan diakhiri dengan penarikan kesimpulan.

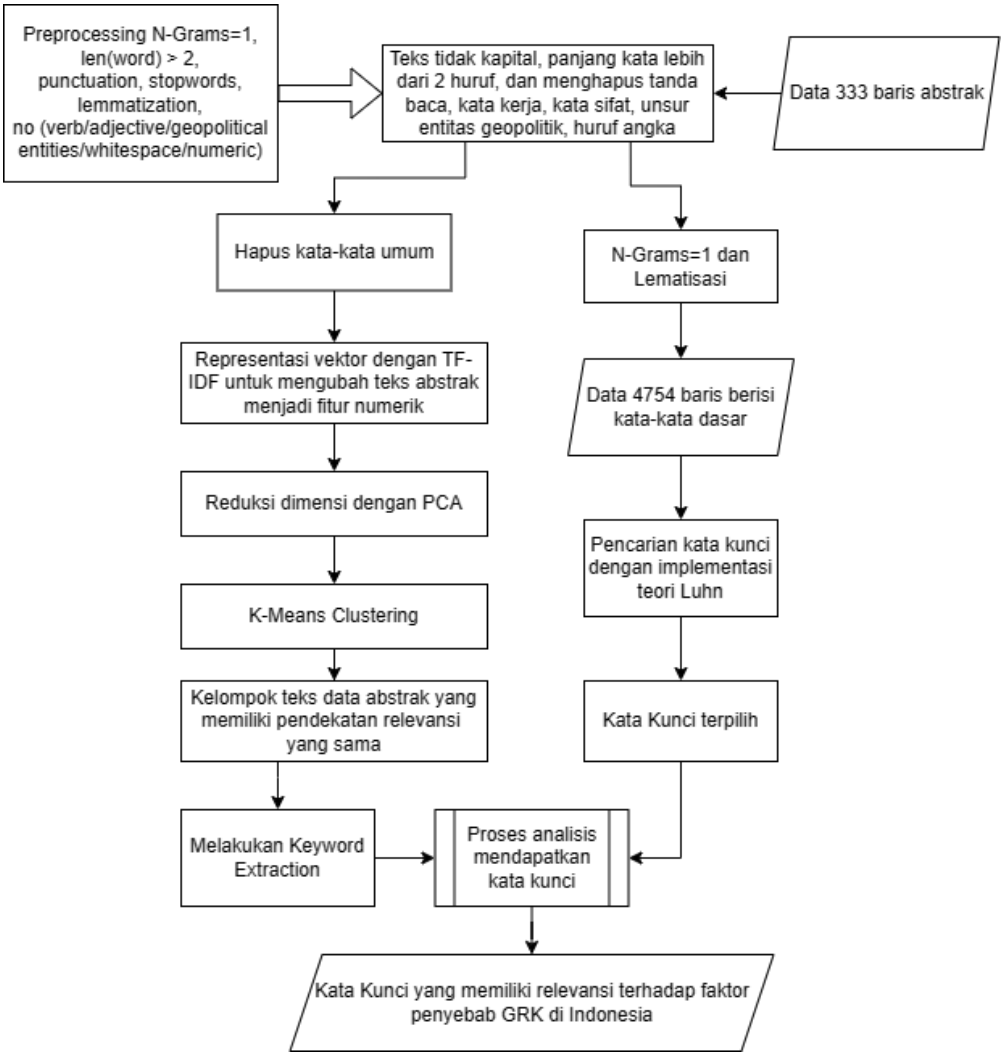

Universitas
Pertamina

BAB IV
HASIL DAN PEMBAHASAN

4.1 *Preprocessing Data*

Data yang telah dikumpulkan berupa 333 dokumen dengan abstrak, pada penelitian ini akan dipakai sehingga berisi 333 baris. Setelah itu, dilakukan suatu *preprocessing data* pada teks tersebut. Pada setelah langkah *preprocessing data* teks tersebut, akan dibagi menjadi dua alur sebagaimana yang telah digambarkan pada Gambar 3.4, yang pertama ialah sebelum dilakukan suatu pengelompokkan abstrak serupa menjadi klaster dan yang alur lainnya sebelum dilakukan pemilihan kata dasar untuk mencari kata kunci yang relevan dengan teori Luhn.

Adapun tahapan dalam *preprocessing data* hingga mendapatkan hasil berupa kata kunci yang dicari digambarkan pada gambar di bawah ini:



Gambar 4.1. Proses mendapatkan kata-kata dasar dari data studi literatur

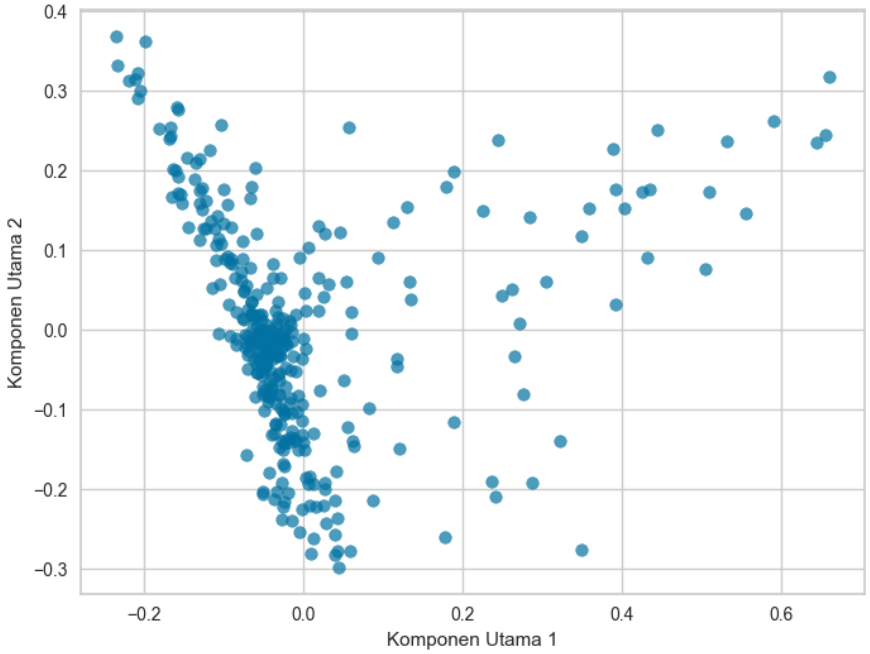
Di bawah ini, merupakan teknik berupa istilah yang digunakan pada saat melakukan *preprocessing data* sebagai berikut:

1. *Lowercase*: Membuat semua teks pada data tersebut menjadi bukan huruf kapital.
2. *Punctuation*: Menghapus semua unsur tanda baca.
3. *Stopword*: Menghapus unsur kata-kata yang tidak bermakna, seperti '*is, a, it, in, he, is*'.
4. N-Grams=1 (Unigram): Mendapatkan data yang berisi dari kumpulan kata-kata dasar.
5. *len(word)* > 2: Mendapatkan kata-kata yang memiliki jumlah huruf lebih dari dua.
6. *Lemmatization*: Memilih satu dari setiap kata dasar yang memiliki bentuk lema yang sama, seperti terdapat kata '*organizing*' dan '*organize*', maka akan dipilih '*organize*'.
7. *No Verb and No Adjective*: Menghapus unsur teks yang mengandung kata kerja dan kata sifat.
8. *No Geopolitical Entities* (GPE): Menghapus semua unsur teks yang mengandung unsur nama-nama negara, kota, atau wilayah.
9. *No Numeric*: Menghapus kata-kata yang memiliki unsur angka saja.
10. *No Whitespace*: Menghapus unsur yang hanya berisi kata spasi.

4.2 Hasil Pengelompokkan Kata Kunci berdasarkan teknik *clustering*

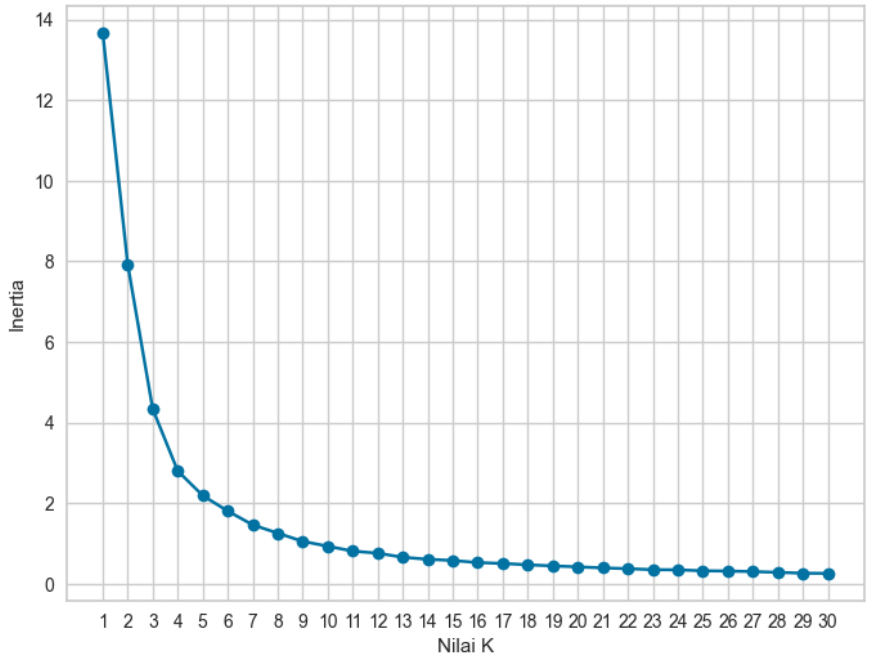
Pada tahapan ini, hasil dari *preprocessing text* sebelumnya, dilakukan suatu proses penghapusan kata-kata umum dengan *stopword* seperti yang dijelaskan pada poin sebelumnya. Kemudian, data yang telah didapatkan tersebut diproses menjadi representasi numerik dari teks abstrak menggunakan metode TF-IDF. TF-IDF digunakan untuk menghitung bobot kata-kata dalam setiap abstrak berdasarkan frekuensi kemunculan kata tersebut di abstrak tertentu dan juga frekuensi kemunculan kata tersebut di semua abstrak. Dengan menggunakan *TfidfVectorizer*, teks abstrak dikonversi menjadi representasi vektor TF-IDF dengan memperhitungkan bobot kata dalam setiap abstrak.

Setelah representasi TF-IDF dibuat, langkah selanjutnya adalah melakukan reduksi dimensi menggunakan PCA. PCA digunakan untuk mengurangi dimensi vektor TF-IDF menjadi dimensi yang lebih rendah, sehingga memungkinkan untuk memvisualisasikan data dalam ruang dua atau tiga dimensi. Dalam implementasinya, digunakan *PCA(n_components=2)* untuk mengambil dua komponen utama dan memvisualisasikan datanya dalam ruang dua dimensi. Komponen utama 1 dan komponen utama 2 pada klasterisasi dengan K-Means merujuk pada hasil reduksi dimensi menggunakan teknik Principal Component Analysis (PCA). Jika nilai komponen utama 1 (nilai x) positif, itu berarti abstrak tersebut cenderung memiliki kontribusi positif dari fitur-fitur yang berkaitan dengan komponen utama 1 tersebut. Sebaliknya, jika nilai komponen utama 1 (nilai x) negatif, itu berarti abstrak tersebut cenderung memiliki kontribusi negatif dari fitur-fitur yang berkaitan dengan komponen utama 1 tersebut.



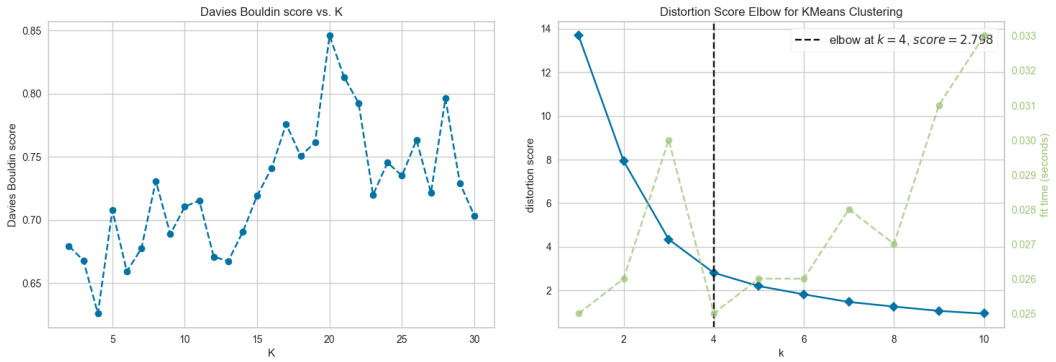
Gambar 4.2. Visualisasi hasil dari PCA

Setelah visualisasi PCA, langkah selanjutnya adalah melakukan K-Means *clustering* untuk mengelompokkan abstrak menjadi beberapa klaster berdasarkan fitur TF-IDF yang telah direduksi menggunakan PCA. Pada implementasi KMeans, dibutuhkan satu parameter yaitu nilai dari K. Untuk menentukan nilai K atau jumlah *num_clusters* yang optimal, dilakukan pencarian dengan menampilkan *Elbow Plot* menggunakan K-Means, dengan uji coba rentang klaster 1-30.



Gambar 4.3. *Elbow Plot* untuk menenentukan nilai klaster dengan K-Means

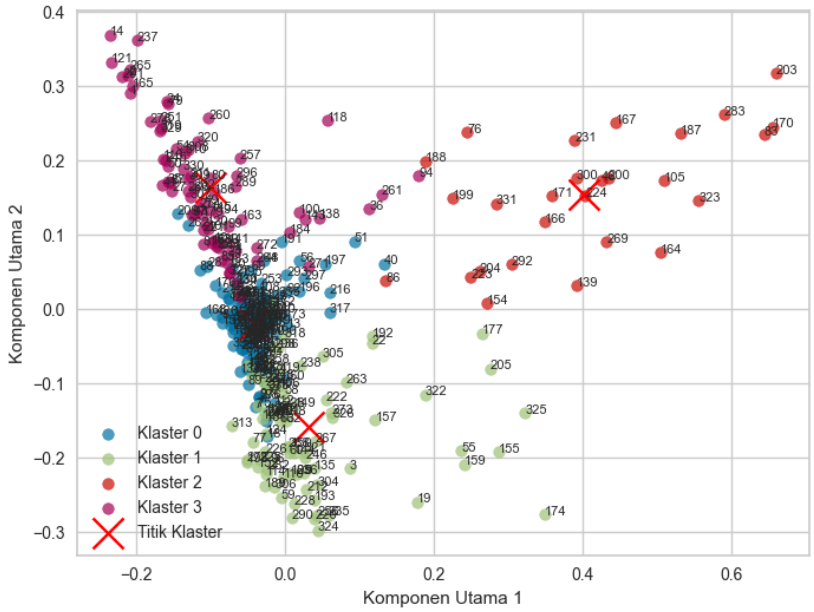
Dalam mempermudah penentuan nilai K yang optimum, dilakukan beberapa metrik evaluasi dengan *Davies-Bouldin index* dan *library ElbowVisualizer*, kedua metrik evaluasi tersebut digunakan untuk mengevaluasi nilai K yang optimal dalam algoritma K-Means. *Davies-Bouldin index* mengukur kualitas pembagian klaster berdasarkan jarak antara klaster yang berdekatan dan dispersi internal setiap klaster [30]. Semakin rendah nilai *Davies-Bouldin index*, semakin baik pembagian klaster dan semakin optimal nilai K yang digunakan. Selain itu, dengan *library ElbowVisualizer* untuk membantu dan memperjelas dari visualisasi *Elbow Plot*, pada saat titik di nilai *inertia* mulai menurun secara lambat atau berhenti menurun drastis, itulah titik yang disarankan sebagai jumlah klaster yang optimal. Hasil dari kedua evaluasi tersebut menunjukkan nilai K=4 yang merupakan nilai K yang optimal, sehingga diterapkan fungsinya yakni *KMeans(n_clusters=4)*.



Gambar 4.4. Metrik Evaluasi nilai K dengan *Davies-Bouldin index* dan *library ElbowVisualizer*

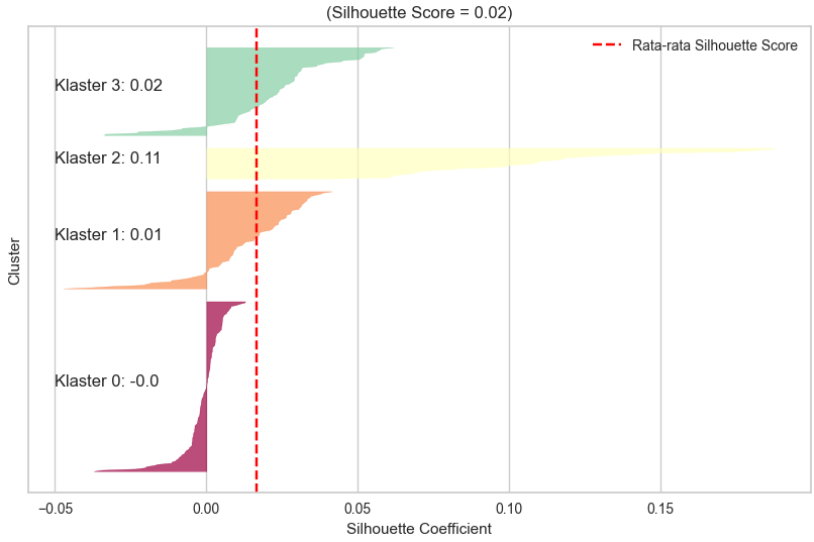
Setelah melakukan K-Means *clustering* dan menentukan jumlah klaster yang optimal berdasarkan *elbow plot* dengan didukung validasi berupa evaluasi metrik, selanjutnya dapat diketahui dalam setiap klaster dengan menggunakan atribut dari objek K-Means untuk menyimpan label klaster pada setiap sampel data abstrak dan mengelompokkannya berdasarkan warna yang berbeda.

Dalam konteks klasterisasi perwarnaan dengan K-Means, komponen utama 1 dan komponen utama 2 adalah dua dimensi baru yang dihasilkan dari PCA. Setiap baris data dalam dataset asli direpresentasikan oleh dua nilai pada komponen utama 1 dan komponen utama 2. Nilai-nilai ini menggambarkan lokasi relatif data dalam ruang dua dimensi yang telah direduksi. Pada saat PCA telah dilakukan, visualisasi data abstrak dalam ruang dua dimensi dengan *scatter plot* menunjukkan bahwa setiap titik pada *scatter plot* merepresentasikan satu abstrak, dan dapat dilihat berupa pola dan struktur data sebelum melakukan *clustering*.



Gambar 4.5. Klasterisasi Pewarnaan dengan K-Means

Selain itu, untuk mempermudah dalam melihat klaster mana yang lebih baik, dilakukanlah *Silhouette plot* yang merupakan metode untuk mengukur seberapa baik setiap data abstrak dikelompokkan oleh K-Means. Plot ini menunjukkan nilai *silhouette score* untuk setiap titik klaster pada data abstrak, nilai *silhouette score* mengukur seberapa mirip data abstrak dengan kelompoknya sendiri dibandingkan dengan kelompok lain. *Silhouette plot* juga membantu mengidentifikasi apakah ada kelompok yang tidak terdefinisi dengan baik atau apakah data abstrak tergolong dalam kelompok yang sesuai.



Gambar 4.6. *Silhouette Plot* untuk K-Means *Clustering*

Dari hasil tersebut, terdapat klaster yang memiliki warna yang menjulang jauh melebihi rata-rata dari *Silhouette Score* yang telah didapatkan senilai 0.02 tersebut, yang menandakan bahwa klaster tersebut memiliki *Silhouette Coefficient* yang tinggi dibandingkan dengan klaster lainnya. Klaster

dengan *Silhouette Coefficient* yang tinggi menunjukkan bahwa titik-titik data dalam klaster tersebut lebih dekat satu sama lain dan lebih terpisah dari titik-titik data di klaster lain. Hal ini, menunjukkan klaster tersebut memiliki titik-titik data yang sangat serupa dan kohesif. Dalam konteks analisis *clustering*, hal ini dapat diartikan sebagai kelompok yang lebih homogen atau memiliki karakteristik yang sama.

Kemudian, pada tahap akhir, dilakukan suatu *keyword extraction* untuk menentukan kata-kata kunci yang paling relevan dalam setiap klaster. Untuk setiap klaster, dilakukan *keyword extraction* masih dengan metode TF-IDF. TF-IDF digunakan untuk menentukan kata-kata kunci yang paling relevan dan memiliki bobot tertinggi dalam klaster tersebut. Setelah mendapatkan kata-kata kunci untuk setiap klaster, akan dihasilkan berupa kata-kata kunci dari proses *keyword extraction*. Berikut ini, 20 kata-kata kunci pada setiap klasternya:

a. ***Cluster 0***: *emissions, indonesia, greenhouse, carbon, ghg, study, emission, gas, climate, change, co2, results, development, research, countries, data, production, impact, food, impacts*

b. ***Cluster 1***: *emissions, forest, soil, peat, rice, ch4, indonesia, greenhouse, land, gas, study, carbon, water, ghg, n2o, emission, area, peatlands, management, palm*

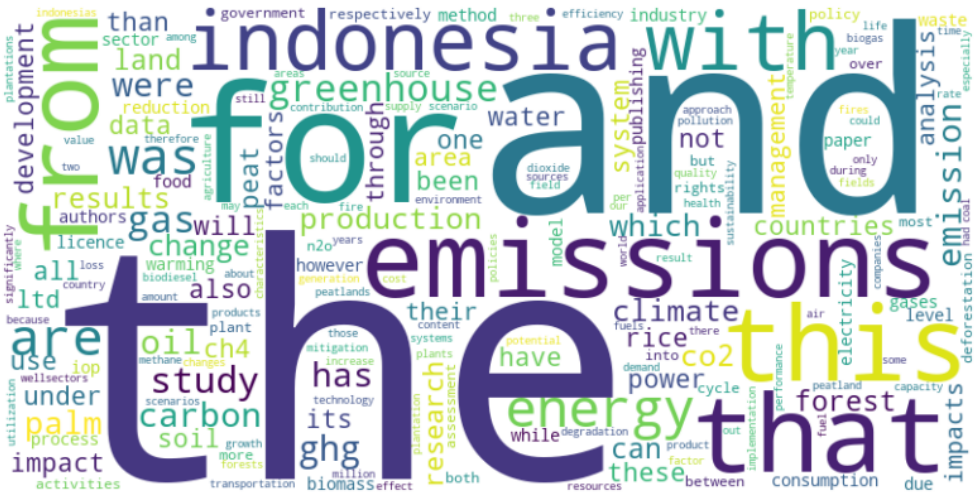
c. ***Cluster 2***: *oil, palm, indonesia, production, biodiesel, emissions, industry, gas, land, greenhouse, study, products, impact, supply, ghg, change, impacts, ispo, cpo, rspo*

d. ***Cluster 3***: *energy, power, indonesia, gas, emissions, electricity, production, greenhouse, co2, study, generation, emission, plant, consumption, ghg, coal, system, use, biogas, efficiency*

4.3 Hasil Pemilihan Kata Kunci berdasarkan Teori Luhn

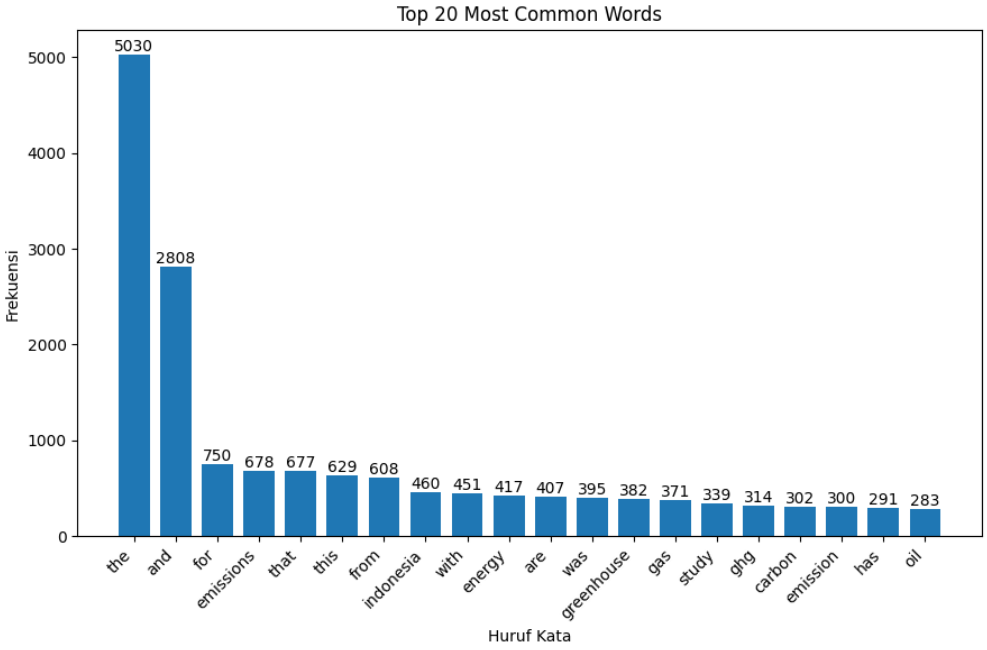
Pada tahapan memilih kata kunci, berdasarkan teori Luhn, data yang telah didapatkan dari hasil *preprocessing data* sebelumnya di bagian 4.1, dilakukan suatu pemecahan kata-kata dasar dengan N-Grams = 1 (Unigram) dan melakukan teknik lematisasi yang telah dijelaskan juga di bagian 4.1.

Setelah dilakukan proses dengan teknik tersebut, didapatkan sebanyak 4754 baris data berupa kata-kata dasar berdasarkan teknik *preprocessing* yang diterapkan sebelumnya. Kemudian untuk melihat suatu frekuensi dari masing-masing kata dasar, terdapat visualisasi dengan *Word cloud* untuk menunjukkan representasi visual dari data teks yang menggambarkan frekuensi kata-kata dalam teks tersebut [31]. Dalam *word cloud*, kata-kata yang muncul lebih sering akan ditampilkan dengan ukuran yang lebih besar, sedangkan kata-kata yang muncul lebih jarang akan ditampilkan dengan ukuran yang lebih kecil.



Gambar 4.7. Representasi visual frekuensi data dengan *word cloud*

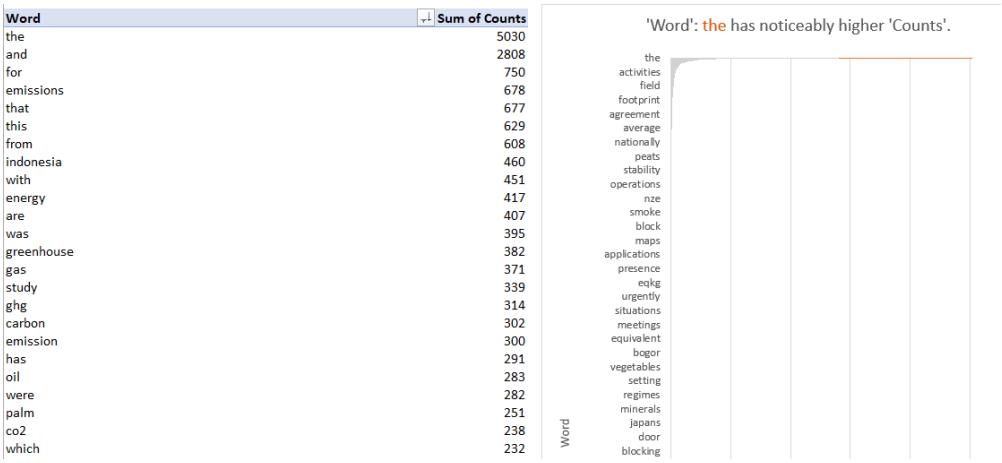
Terdapat juga visualisasi untuk melihat 20 kata yang memiliki frekuensi terbesar pada data tersebut.



Gambar 4.8. Dua puluh kata dasar dengan frekuensi terbesar

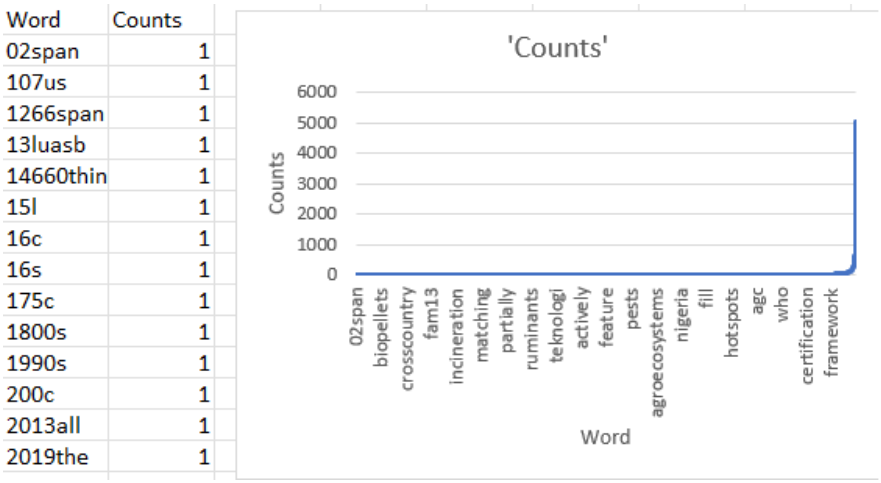
Namun, yang diproses seleksi teori Luhn bukanlah kata-kata yang memiliki frekuensi terbesar, semua data sebanyak 4754 kata sebelumnya, akan dipakai pada tahapan teori Luhn. Adapun bagian-bagian kata yang memiliki frekuensi terbesar dan frekuensi terkecil akan dibuang, karena berdasarkan teori Luhn porsi pada rentang tersebut merupakan istilah-istilah yang kurang baik dalam memilih kata kunci yang relevan secara optimal.

Pada proses seleksi tersebut, dilakukan analisis secara manual dengan melihat kata-kata dasar berdasarkan relevansi yang ingin dipilih.



Word	Sum of Counts
the	5030
and	2808
for	750
emissions	678
that	677
this	629
from	608
indonesia	460
with	451
energy	417
are	407
was	395
greenhouse	382
gas	371
study	339
ghg	314
carbon	302
emission	300
has	291
oil	283
were	282
palm	251
co2	238
which	232

Gambar 4.9. *Preview* kata-kata dengan frekuensi terbesar



Word	Counts
02span	1
107us	1
1266span	1
13luasb	1
14660thin	1
15l	1
16c	1
16s	1
175c	1
1800s	1
1990s	1
200c	1
2013all	1
2019the	1

Gambar 4.10. *Preview* kata-kata dengan frekuensi terendah

Agar dalam seleksi kata kunci tersebut, setelah dilakukan suatu analisis secara mendalam dengan melihat kata-kata dasar berdasarkan relevansi, bobot frekuensi, dan juga bagian interval yang cukup baik sebagai pemilihan kata kunci, agar lebih presisi, maka porsi besaran pada kata-kata di rentang frekuensi terbesar dan frekuensi terkecil akan dibuang sebesar 10% dari keseluruhan jumlah data yang berarti dihapus 475 kata di bagian rentang frekuensi terbesar dan rentang frekuensi terendah, sehingga tersisa 3804 baris data.

Setelah itu, membuang kata-kata dasar yang memiliki frekuensi lebih dari atau sama dengan 10 sehingga tersisa 169 baris data. Kemudian baru dilakukan rentang mana yang akan dipilih sebagai *upper cut* dan *lower cut* dari data tersebut. Setelah dilakukan analisis kembali, dipilih rentang sebesar 15% untuk dijadikan sebagai kata kunci yang terpilih. Dari hasil kata kunci dipilihlah sebanyak 25 kata yang memiliki relevansi berdasarkan hasil analisis secara manual dan penerapan teori Luhn.

137	central	14	
138	challenges	14	
139	context	14	LOWER CUT
140	**contributions**	14	
141	**control**	14	
142	**criteria**	14	
143	**crops**	14	
144	**damage**	14	
145	**difference**	14	
146	**diversity**	14	
147	**drainage**	14	
148	**future**	14	
149	**generally**	14	
150	**governments**	14	**SELECTED KEYWORDS (15%) [25 Words]**
151	**green**	14	
152	**half**	14	
153	**household**	14	
154	**ice**	14	
155	**petroleum**	14	
156	**temperatures**	14	
157	**priority**	14	
158	**recycling**	14	
159	**risk**	14	
160	**root**	14	
161	**simulation**	14	
162	**soils**	14	
163	**speed**	14	
164	**states**	14	
165	pome	14	UPPER CUT
166	tool	14	
167	usage	14	

Gambar 4.11. Hasil Kata Kunci Berdasarkan Hasil Analisis Manual dan Implementasi Teori Luhn

Hasil dari 25 kata tersebut yaitu '*contributions, control, criteria, crops, damage, difference, diversity, drainage, future, generally, governments, green, half, household, ice, petroleum, temperatures, priority, recycling, risk, root, simulation, soils, speed, states*'

4.4 Hasil Kata Kunci yang Didapatkan Berdasarkan Teknik *Clustering* dan Seleksi Teori Luhn

Telah dihasilkan berupa kata-kata kunci dari hasil teknik *Clustering* dan seleksi menggunakan Teori Luhn, yaitu:

Tabel 4.1. Hasil Kata Kunci yang Terpilih Berdasarkan Teknik *Clustering* dan Implementasi Teori Luhn

Kata Kunci Terpilih	
Cluster 0 Keywords	*emissions, indonesia, greenhouse, carbon, ghg, study, emission, gas, climate, change, co2, results, development, research, countries, data, production, impact, food, impacts*
Cluster 1 Keywords	*emissions, forest, soil, peat, rice, ch4, indonesia, greenhouse, land, gas, study, carbon, water, ghg, n2o, emission, area, peatlands, management, palm*
Cluster 2 Keywords	*oil, palm, emissions, indonesia, production, gas, greenhouse, biodiesel, industry, ghg, study, soil, land, forest, peat, plantations, n2o, carbon, co2, impact*
Cluster 3 Keywords	*energy, power, indonesia, gas, emissions, electricity, production, greenhouse, co2, study, generation, emission, plant, consumption, ghg, coal, system, use, biogas, efficiency*
25 Kata dengan Teori Luhn	*contributions, control, criteria, crops, damage, difference, diversity, drainage, future, generally, governments, green, half, household, ice, petroleum, temperatures, priority, recycling, risk, root, simulation, soils, speed, states*

Jika dilakukan suatu pemilihan kata-kata kunci dari kedua hasil proses tersebut, berdasarkan sumber dari data studi literatur, aspek dan faktor penyebab dari Gas Rumah Kaca yang ada di Indonesia, maka dari hasil tersebut bisa dijadikan sebagai landasan secara efektif untuk melihat kondisi saat ini dengan memanfaatkan studi literatur 5 tahun terakhir yang membahas terkait subjek atau urgensi perihal GRK di Indonesia. Contohnya seperti *electricity consumption, countries emission, biodiesel production, energy use* dan kata-kata yang memiliki keterkaitan antara sesama kata-kata kunci yang bisa digabungkan menjadi frasa seperti contoh tersebut, adapun pada penelitian ini terdapat berbagai banyak frasa lainnya yang dapat dibentuk dari hasil klasterisasi dan juga kata-kata kunci dari teori Luhn (lihat halaman Lampiran B). Oleh karena itu, dari hasil penelitian ini, bisa menjadi salah satu perhatian bahwa dengan memanfaatkan teknologi saat ini seperti *Artificial Intellegence* dan juga analisis data yang didukung juga berdasarkan teori, dapat membantu dalam analisis tersebut.


Universitas
Pertamina

BAB V
KESIMPULAN DAN SARAN

5.1 Kesimpulan

Dari hasil analisis data teks menggunakan pengolahan dan teknik dengan *text mining*, *clustering*, dan penyeleksian menggunakan teori Luhn pada data abstrak dari studi literatur lima tahun terakhir, telah didapatkan kata-kata kunci yang memiliki relevan sebagai faktor penyebab Gas Rumah Kaca (GRK) di Indonesia. Hasil analisis ini dapat memberikan informasi yang bermanfaat terkait permasalahan GRK saat ini dan dapat menjadi landasan untuk langkah-langkah penanganan dan mitigasi lebih lanjut terhadap dampak GRK di Indonesia.

5.2 Saran

Untuk penelitian selanjutnya, disarankan agar dapat memperluas sumber data studi literatur dengan mencakup lebih banyak publikasi terkini dan diversifikasi sumber literatur, termasuk jurnal ilmiah, laporan pemerintah, dan publikasi akademis lainnya. Penggunaan sumber data yang lebih luas akan memberikan gambaran yang lebih komprehensif dan akurat terkait faktor penyebab GRK di Indonesia.

Selanjutnya, dapat dilakukan kajian lebih lanjut dengan mempertimbangkan aspek waktu, geografis, dan perubahan iklim, sehingga dapat memberikan pemahaman yang lebih holistik tentang dampak GRK di Indonesia. Dengan menggunakan analisis data yang lebih maju dan inklusif, pemanfaatan teknologi tepat, hasil yang diharapkan dapat memberikan kontribusi yang lebih signifikan dalam pengembangan strategi dan kebijakan mitigasi GRK yang lebih efektif dan berkelanjutan di Indonesia.


Universitas
Pertamina

DAFTAR PUSTAKA

[1] M. A. Rizaty, "Emisi Gas Rumah Kaca Indonesia Diproyeksi Terus Naik hingga 2030," *DataIndonesia.Id*, Oct. 2022. [Online]. Available: https://dataindonesia.id/varia/detail/emisi-gas-rumah-kaca-indonesia-diproyeksi-terus-naik-hingga-2030

[2] G. A. Kristanto and W. Koven, "Estimating greenhouse gas emissions from municipal solid waste management in Depok, Indonesia," *City and Environment Interactions*, vol. 4, p. 100027, Dec. 2019. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2590252020300088

[3] D. G. K. Ketaren, "PERANAN KAWASAN MANGROVE DALAM PENURUNAN EMISI GAS RUMAH KACA DI INDONESIA," *Jurnal Kelautan dan Perikanan Terapan (JKPT)*, vol. 1, p. 73, Jan. 2023. [Online]. Available: http://ejournal-balitbang.kkp.go.id/index.php/jkpt/article/view/12050

[4] J. Li, M. Irfan, S. Samad, B. Ali, Y. Zhang, D. Badulescu, and A. Badulescu, "The Relationship between Energy Consumption, CO2 Emissions, Economic Growth, and Health Indicators," *International Journal of Environmental Research and Public Health*, vol. 20, no. 3, p. 2325, Jan. 2023. [Online]. Available: https://www.mdpi.com/1660-4601/20/3/2325

[5] M. Fajri, I. Adi Nugroho, and T. Tri Hendro Atmoko Utomo, "Standar lhk, penunjang target pencapain folu net sink 2030," *STANDAR: Better Standard Better Living*, vol. 2, no. 3, p. 14–18, Mei 2023. [Online]. Available: http://majalah.bsilhk.menlhk.go.id/index.php/STANDAR/article/view/124

[6] S. P. Tiwari, S. Prasad, and M. Thushara, "Machine learning for translating pseudocode to python: A comprehensive review," in *2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS)*, May 2023, pp. 274–280.

[7] S. Salloum, M. Al-Emran, A. Monem, and K. Shaalan, *Using Text Mining Techniques for Extracting Information from Research Articles*, 01 2018, pp. 373–397.

[8] H. P. Luhn, "The Automatic Creation of Literature Abstracts," *IBM Journal of Research and Development*, vol. 2, no. 2, pp. 159–165, Apr. 1958. [Online]. Available: http://ieeexplore.ieee.org/document/5392672/

[9] Computer Education Department, ACLC College of Butuan, Butuan City, Philippines, V. Z. V. Singco, J. C. Trillo, C. C. Abalorio, J. C. M. Bustillo, J. T. Bojocan, and M. C. Elape, "OCR-based Hybrid Image Text Summarizer using Luhn Algorithm with FinetuneTransformer Modelsfor Long Document," *International Journal of Emerging Technology and Advanced Engineering*, vol. 13, no. 2, pp. 47–56, Feb. 2023. [Online]. Available: https://www.ijetae.com/files/Volume13Issue2/IJETAE_0223_07.pdf

[10] A. Rahmah, H. B. Santoso, and Z. A. Hasibuan, "Critical Review of Technology-Enhanced Learning using Automatic Content Analysis," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 1, 2022. [Online]. Available: http://thesai.org/Publications/ViewPaper?Volume=13&Issue=1&Code=IJACSA&SerialNo=48

[11] H. Wang, J. Luo, M. Zhang, and Y. Ling, "The Impact of Transportation Restructuring on the Intensity of Greenhouse Gas Emissions: Empirical Data from China," *International Journal of Environmental Research and Public Health*, vol. 19, no. 19, p. 12960, Oct. 2022. [Online]. Available: https://www.mdpi.com/1660-4601/19/19/12960

[12] E. Purnamasari, S. Sudarno, and H. Hadiyanto, "INVENTARISASI EMISI GAS RUMAH KACA SEKTOR PERTANIAN DI KABUPATEN BOYOLALI," *Universitas Muhammadiyah Surakarta*, Apr. 2019.

[13] R. Pratama, "EFEK RUMAH KACA TERHADAP BUMI," vol. 14, no. 2, 2019.

[14] K. O. Yoro and M. O. Daramola, "CO2 emission sources, greenhouse gases, and the global warming effect," in *Advances in Carbon Capture*. Elsevier, 2020, pp. 3–28. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/B9780128196571000013

[15] K. Musari and F. Sayah, "Green Financing through Green Sukuk in the Fight Against Climate Change: Lessons from Indonesia," 2021, publisher: Unpublished. [Online]. Available: http://rgdoi.net/10.13140/RG.2.2.23804.26245

[16] A. Aqlan, D. M. Bairam, and R. L. Naik, *A Study of Sentiment Analysis: Concepts, Techniques, and Challenges*, 01 2019, pp. 147–162.

[17] S. Chung, S. Moon, J. Kim, J. Kim, S. Lim, and S. Chi, "Comparing natural language processing (nlp) applications in construction and computer science using preferred reporting items for systematic reviews (prisma)," *Automation in Construction*, vol. 154, p. 105020, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0926580523002807

[18] S. Moon, S. Chi, and S.-B. Im, "Automated detection of contractual risk clauses from construction specifications using bidirectional encoder representations from transformers (bert)," *Automation in Construction*, vol. 142, p. 104465, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0926580522003387

[19] M. Nasimuzzaman, A. N. Merag, S. Afroj, M. M. Alam, M. H. K. Mehedi, and A. A. Rasel, "Consumer review analysis using nlp and data mining," in *2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC)*, March 2023, pp. 0426–0430.

[20] Y. Setiawan, D. Gunawan, and R. Efendi, "Feature extraction tf-idf to perform cyberbullying text classification: A literature review and future research direction," in *2022 International Conference on Information Technology Systems and Innovation (ICITSI)*. IEEE, 2022, pp. 283–288.

[21] K. L. Tan, C. P. Lee, and K. M. Lim, "A survey of sentiment analysis: Approaches, datasets, and future research," *Applied Sciences*, vol. 13, no. 7, p. 4550, Apr 2023. [Online]. Available: http://dx.doi.org/10.3390/app13074550

[22] T. Kurita, "Principal component analysis (pca)," *Computer Vision: A Reference Guide*, pp. 1–4, 2019.

[23] J. R. Beattie and F. W. L. Esmonde-White, "Exploration of Principal Component Analysis: Deriving Principal Component Analysis Visually Using Spectra," *Applied Spectroscopy*, vol. 75, no. 4, pp. 361–375, Apr. 2021. [Online]. Available: http://journals.sagepub.com/doi/10.1177/0003702820987847

[24] A. F. Jahwar and A. M. Abdulazeez, "Meta-heuristic algorithms for k-means clustering: A review," *PalArch's Journal of Archaeology of Egypt/Egyptology*, vol. 17, no. 7, pp. 12 002–12 020, 2020.

[25] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," *Information Sciences*, vol. 622, pp. 178–210, Apr. 2023. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0020025522014633

[26] B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering – A systematic literature review," *Information and Software Technology*, vol. 51, no. 1, pp. 7–15, Jan. 2009. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0950584908001390

[27] S. Pizard, F. Acerenza, D. Vallespir, and B. Kitchenham, "Assessing attitudes towards evidence-based software engineering in a government agency," *Information and Software Technology*, vol. 154, p. 107101, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950584922002105

[28] L. S. L. Lai, "CONTENT ANALYSIS OF SOCIAL MEDIA: A GROUNDED THEORY APPROACH," *Social Media Content Analysis*, vol. 16, no. 2, 2015.

[29] A. Yunita, H. B. Santoso, and Z. A. Hasibuan, "Finding Contributing Factors of Students' Academic Achievement Using Quantitative and Qualitative Analyses-Based Information Extraction," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 17, no. 16, pp. 108–125, Aug. 2022. [Online]. Available: https://online-journals.org/index.php/i-jet/article/view/31945

[30] S. Lee, G. Hahn, J. Hecker, S. M. Lutz, K. Mullin, A. D. N. I. (ADNI), W. Hide, L. Bertram, D. L. DeMeo, R. E. Tanzi, C. Lange, and D. Prokopenko, "A comparison between similarity matrices for principal component analysis to assess population stratification in sequenced genetic data sets," *Briefings in Bioinformatics*, vol. 24, no. 1, p. bbac611, 12 2022. [Online]. Available: https://doi.org/10.1093/bib/bbac611

[31] R. M. M. Hicke, M. Goenka, and E. Alexander, "Word Clouds in the Wild," in *2022 IEEE 7th Workshop on Visualization for the Digital Humanities (VIS4DH)*. Oklahoma City, OK, USA: IEEE, Oct. 2022, pp. 43–48. [Online]. Available: https://ieeexplore.ieee.org/document/9974812/

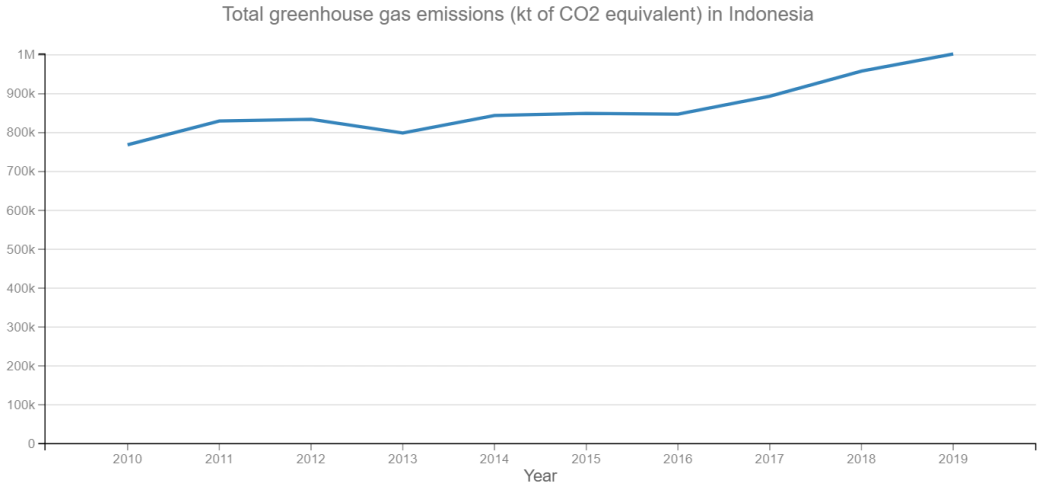
[32] World Bank, "World development indicators databank," The World Bank, Washington, DC, Tech. Rep., Jul. 2023. [Online]. Available: https://databank.worldbank.org/source/world-development-indicators


Universitas
Pertamina

LAMPIRAN A

Lampiran 1. Tren Total Emisi GRK Indonesia

Data yang diambil dari *World Development Indicators* oleh World Bank [32], menunjukkan *total greenhouse gas emissions (kt of CO2 equivalent)* di Indonesia tahun 2010-2019. Mengindikasikan bahwa tren dari emisi yang ditimbulkan oleh GRK di Indonesia semakin meningkat secara signifikan, pada tahun 2019 dihasilkan sebesar 1.002.369,0 kt of CO2.



Gambar 1.1. Total emisi GRK di Indonesia 2010-2019

LAMPIRAN B

Lampiran 2. Frasa dari Kata Kunci yang Didapatkan

Hasil dari kata-kata kunci dengan adanya hasil dari klasterisasi dan teori Luhn, berikut ini berbagai frasa dengan rentang 1-2 kata setiap frasanya. Untuk membantu hal tersebut, digunakan *library Keybert* yang mampu mengidentifikasi kata-kata yang mencerminkan makna dari teks yang dipakai, dalam hal ini dipakai teks dari hasil klasterisasi.

Gambar 2.1. Hasil frasa yang dapat dibentuk (1)

adoption', 0.3984), ('effects strategy', 0.3921), ('effect carbon', 0.3867)], [('waste management', 0.6437), ('waste
recycling', 0.6412), ('businesses waste', 0.6363), ('recycling businesses', 0.6357), ('plastic waste', 0.6201)],
[('drought assessment', 0.5698), ('data drought', 0.5396), ('drought severity', 0.5367), ('severity drought', 0.5252),
('drought prediction', 0.5235)], [('impacts climate', 0.5805), ('indonesia risks', 0.5504), ('community indonesia',
0.5404), ('contribution climate', 0.5313), ('change indonesias', 0.5169)], [('coal power', 0.6455), ('effect coal',
0.6045), ('characteristics coals', 0.5856), ('performance coals', 0.5848), ('coals coals', 0.5804)], [('waste
composition', 0.7051), ('waste type', 0.658), ('waste dominantly', 0.6355), ('landfill recycling', 0.6244), ('bakung
landfill', 0.6152)], [('moratorium deforestation', 0.4981), ('deforestation forest', 0.4872), ('emissions ecosystems',
0.4865), ('conservation policies', 0.4862), ('conservation policys', 0.4794)], [('tuna longlines', 0.5664), ('tuna
longline', 0.5662), ('cfp tuna', 0.5543), ('tuna fisheries', 0.5326), ('lines tuna', 0.5218)], [('ghg emissions',
0.6011), ('determinants ghg', 0.5631), ('ghg emission', 0.5411), ('emissions reasons', 0.5286), ('characteristics ghg',
0.5176)], [('co2 atmosphere', 0.6868), ('emissions co2', 0.6555), ('co2 emission', 0.6336), ('emission co2', 0.6251),
('study co2', 0.6055)], [('contributors emissions', 0.5129), ('emissions indonesia', 0.5014), ('readiness indonesia',
0.4904), ('fuel mobility', 0.4775), ('emissions goal', 0.4755)], [('oil indonesias', 0.5252), ('indonesia level',
0.4273), ('oilseeds sector', 0.4229), ('palm oil', 0.4218), ('energy directive', 0.419)], [('emissions countries',
0.648), ('emission economies', 0.6292), ('emissions taxation', 0.583), ('carbon emissions', 0.5566), ('carbon
neutrality', 0.5544)], [('emissions pattern', 0.5275), ('emissions landsat', 0.514), ('emission patterns', 0.4741),
('carbon emissions', 0.4667), ('relationship carbon', 0.4594)], [('emissions logistics', 0.5259), ('emissions
transportation', 0.5242), ('emissions impact', 0.5011), ('indonesia economy', 0.4842), ('emissions increase', 0.4778)],
[('indonesia variables', 0.4822), ('environment indonesia', 0.4655), ('quality indonesia', 0.4569), ('indonesia grdp',
0.434), ('variables research', 0.4193)], [('sustainability impacts', 0.5341), ('sustainability assessment', 0.5164),
('assessment sustainability', 0.492), ('sustainability sdg', 0.4611), ('mining project', 0.4426)], [('emissions value',
0.5003), ('analysis greenhouse', 0.4805), ('tonneday waste', 0.4766), ('emission factors', 0.4757), ('emissions co2eqt',
0.468)], [('emissions study', 0.552), ('emissions reporting', 0.5289), ('emissions operations', 0.5206), ('companys
carbon', 0.5113), ('carbon emissions', 0.4994)], [('dependence fuels', 0.6148), ('emissions research', 0.5808), ('gas
emissions', 0.5762), ('fuels earths', 0.5733), ('fossil fuels', 0.572)], [('policies adoption', 0.5074), ('indonesias
development', 0.4875), ('emissions market', 0.4853), ('strategies indonesias', 0.4801), ('development adoption',
0.4583)], [('emissions transportation', 0.5385), ('logistics distribution', 0.5195), ('logistics route', 0.5172),
('logistics multidepot', 0.5035), ('study logistics', 0.4933)], [('environment construction', 0.6007), ('construction
materials', 0.5438), ('construction environment', 0.5367), ('construction tests', 0.5289), ('ecofriendly construction',
0.5152)], [('plastic production', 0.5687), ('plastic debris', 0.5673), ('climate waste', 0.5617), ('sea plastic',
0.5612), ('debris policy', 0.5564)], [('sba bioenergy', 0.612), ('resources biomass', 0.5532), ('biomass waste', 0.5502),
('production bioenergy', 0.517), ('biomass', 0.5145)], [('climate study', 0.4841), ('studies warming', 0.4742), ('topics
science', 0.4576), ('climate change', 0.4508), ('warming climate', 0.4236)], [('emissions batteries', 0.6197), ('battery
vehicles', 0.5873), ('emissions lithium', 0.5779), ('ghg emissions', 0.542), ('battery lib', 0.5185)], [('indonesia
frameworks', 0.4882), ('ghg inventories', 0.447), ('implementation ghg', 0.4293), ('frameworks contexts', 0.4088),
('framework implementation', 0.4087)], [('cultivation training', 0.5744), ('indonesia agriculture', 0.477), ('lands
greenhouse', 0.4639), ('farmers practices', 0.4602), ('harvests processing', 0.4328)], [('coffee production', 0.58),
('wastes coffee', 0.5358), ('coffee potential', 0.5216), ('coffee processing', 0.5168), ('ground coffee', 0.4775)],
[('emissions ghg', 0.6045), ('ghg emissions', 0.5888), ('emissions inventory', 0.5819), ('inventory emissions', 0.5796),
('emissions profile', 0.5665)], [('sea emissions', 0.7049), ('emissions climate', 0.6279), ('emissions sector', 0.612),
('emissions inventory', 0.6076), ('emissions period', 0.5989)], [('indonesia carbon', 0.6325), ('emissions impact',
0.5962), ('emissions climate', 0.5926), ('impact indonesias', 0.583), ('indonesias household', 0.5603)], [('emissions
indonesia', 0.6446), ('greenhouse gas', 0.5669), ('process greenhouse', 0.548), ('plan greenhouse', 0.5317), ('provinces
indonesia', 0.5008)], [('emission commitment', 0.5238), ('evs emissions', 0.5125), ('ghg emission', 0.4456),
('emissions', 0.4451), ('emissions atmosphere', 0.442)], [('indonesia climate', 0.5543), ('ghg indonesia', 0.4693),
('indonesia countries', 0.4586), ('indonesia data', 0.4574), ('climate change', 0.4503)], [('emissions countries',
0.6552), ('emissions impact', 0.6134), ('emissions additionally', 0.5761), ('emissions models', 0.5739), ('nations
emission', 0.5732)], [('landfilling msw', 0.6527), ('landfill msw', 0.6362), ('waste msw', 0.636), ('emissions landfill',
0.6034), ('landfill climate', 0.5959)], [('greenhouse gases', 0.6398), ('gas photosynthesis', 0.6277), ('tree
photosynthesis', 0.6192), ('photosynthesis research', 0.6092), ('atmosphere trees', 0.5921)], [('sugarcane monoculture',
0.4919), ('impacts monoculture', 0.4909), ('indonesia agroecosystem', 0.4902), ('effects monoculture', 0.4787),
('functions agrobiodiversity', 0.4594)], [('warming issue', 0.5773), ('climate change', 0.5773), ('impact climate',
0.5535), ('warming impact', 0.539), ('ghgs climate', 0.5201)], [('biodiesel fuel', 0.7494), ('analysis biodiesel',
0.7138), ('study biodiesel', 0.7001), ('biodiesel requirements', 0.6916), ('biodiesel gas', 0.6902)], [('climate change',
0.5648), ('reason islands', 0.518), ('climate agreement', 0.5054), ('islands areas', 0.5035), ('warming climate',
0.5028)], [('greenhouse gas', 0.5867), ('ghg emissions', 0.5858), ('emissions results', 0.5827), ('emissions combustion',
0.567), ('biomass loss', 0.5607)], [('combustion fuels', 0.6613), ('fuels gasification', 0.6568), ('fuels form', 0.6445),
('gasification biomass', 0.6242), ('gas fuel', 0.5936)], [('ghg inventories', 0.531), ('categories ghg', 0.4974), ('ghg
inventory', 0.4845), ('sector ghg', 0.4715), ('countries categories', 0.4436)], [('eucalyptus adaptability', 0.6454),
('biomass challenge', 0.6049), ('forest management', 0.6007), ('tree biomass', 0.5845), ('forests stability', 0.5709)],
[('scenario countries', 0.5114), ('climate policies', 0.4818), ('countries health', 0.4705), ('cobenefits scenarios',
0.4291), ('health cobenefits', 0.4291)], [('emissions indonesia', 0.6043), ('emissions urbanization', 0.6025),
('emissions sector', 0.5878), ('sector emissions', 0.5803), ('growth emissions', 0.5512)], [('water policies', 0.5918),
('water policy', 0.5886), ('water management', 0.5333), ('government indonesia', 0.5238), ('implementation policies',
0.5118)], [('greenhouse gases', 0.5666), ('climate events', 0.5583), ('ocean warming', 0.5544), ('gases earths', 0.5428),
('impacts climate', 0.5334)], [('processes urbanization', 0.6106), ('urbanization process', 0.5866), ('periurbanization
processes', 0.5386), ('influences urbanization', 0.5266), ('process cities', 0.5024)], [('emissions development',
0.5434), ('government carbon', 0.5209), ('carbon development', 0.5089), ('sectors greenhouse', 0.4814), ('growth
emissions', 0.4778)], [('contributors carbon', 0.6028), ('carbon development', 0.5303), ('ghg contributors', 0.5192),
('sustainability world', 0.4999), ('greenhouse gas', 0.4855)], [('indonesia forestry', 0.5998), ('forestry activities',
0.5623), ('production forest', 0.5385), ('emissions emissions', 0.5326), ('emission carbon', 0.5315)], [('deficiencies
aquaculture', 0.6551), ('impacts aquaculture', 0.6139), ('aquaculture micronutrients', 0.5578), ('aquaculture
strategies', 0.537), ('aquaculture systems', 0.5364)], [('regulation indonesias', 0.5302), ('indonesias greenhouse',
0.5209), ('emission ccs', 0.483), ('contributor carbon', 0.4808), ('carbon capture', 0.4653)], [('emission brunei',
0.6267), ('emissions reasons', 0.5791), ('emission countries', 0.5707), ('singapore analysis', 0.5418), ('emission
philippines', 0.527)], [('emissions logistics', 0.6028), ('logistics carbon', 0.5944), ('emissions policy', 0.5252),
('strategies carbon', 0.5191), ('carbon footprint', 0.5067)], [('commitments deforestation', 0.6087), ('governance
scales', 0.5556), ('role governance', 0.5381), ('jurisdictions commitment', 0.5363), ('deforestation target', 0.5359)],
[('waste business', 0.506), ('wastepaper groundwater', 0.4926), ('waste recycling', 0.461), ('wastepaper wastepaper',
0.446), ('method factor', 0.4402)], [('energy sustainability', 0.4773), ('energy diversification', 0.4733), ('development
energy', 0.4571), ('sustainability greenhouse', 0.4206), ('diversification energy', 0.4152)], [('consumption indonesia',
0.5045), ('indonesia use', 0.4785), ('development indonesia', 0.4638), ('oil dependency', 0.4324), ('oil consumption',
0.4167)], [('plants indonesia', 0.5396), ('distribution indonesia', 0.5031), ('conservation indonesia', 0.4956), ('plant
diversity', 0.4895), ('areas climate', 0.4877)], [('impacts indonesias', 0.5992), ('indonesias capital', 0.5308),
('indonesias greenhouse', 0.5304), ('borneo worlds', 0.5263), ('borneo indonesia', 0.5209)], [('cement industry',
0.6366), ('emissions cement', 0.6149), ('manufacturing cement', 0.5738), ('better cement', 0.516), ('cement', 0.4916)],
[('narratives impact', 0.5021), ('narratives media', 0.4769), ('narratives denialist', 0.454), ('denialist narratives',
0.4403), ('narratives heavily', 0.4373)], [('biofuels attention', 0.6167), ('consumption biofuels', 0.6024), ('biofuels',
0.5533), ('tariffs greenhouse', 0.5355), ('greenhouse gas', 0.5236)], [('indonesias ghg', 0.5393), ('ghg emission',
0.528), ('gases ghg', 0.5058), ('industry emission', 0.4881), ('methodology ghg', 0.4803)], [('indonesia biomass',
0.6881), ('biomass ethylene', 0.6404), ('ethylene biomass', 0.6345), ('biomass feedstock', 0.566), ('biomass material',
0.5652)], [('farms cows', 0.6194), ('cow climate', 0.6152), ('milk yields', 0.6105), ('cow farms', 0.6077), ('farms
milkcowy', 0.6015)], [('extinctions million', 0.5454), ('life extinctions', 0.5399), ('mass extinction', 0.5081),
('extinctions extinctions', 0.505), ('countries greenhouse', 0.5018)], [('climate variability', 0.5814), ('changes
climate', 0.5517), ('ocean changes', 0.5275), ('iod climate', 0.5245), ('role climate', 0.4954)], [('emission livestock',
0.7099), ('emission sheep', 0.6082), ('livestock sector', 0.6026), ('cattle ch4year', 0.5657), ('n2o livestock',
0.5622)], [('contributors pollution', 0.4966), ('pollution development', 0.4653), ('exposure profitability', 0.4624),
('pollution carbon', 0.4602), ('participation pollution', 0.4434)], [('emissions indonesia', 0.6631), ('emissions
impact', 0.5224), ('indonesia need', 0.503), ('emissions', 0.4756), ('gas emissions', 0.4663)], [('emissions road',
0.5745), ('infrastructure ghg', 0.5538), ('asphalt ghg', 0.5455), ('construction roads', 0.5229), ('road construction',
0.4981)], [('indonesias emission', 0.6664), ('emissions relationship', 0.6625), ('emission policies', 0.6163), ('carbon
emissions', 0.6116), ('ghg emissions', 0.5899)], [('efficiency sustainability', 0.6118), ('sustainability fruit',
0.5416), ('factor productivity', 0.5095), ('sustainability', 0.5015), ('production practices', 0.497)], [('sectors
emission', 0.5294), ('emission policy', 0.5279), ('emission pollution', 0.4897), ('emission issues', 0.4493),
('environment citys', 0.4423)], [('criteria questionnaire', 0.3999), ('plastic shopping', 0.3891), ('bags life', 0.371),
('bags sampling', 0.366), ('plastic bags', 0.3453)], [('greenhouse gas', 0.5668), ('carbon dioxide', 0.562), ('carbon
emission', 0.5588), ('evaluation greenhouse', 0.5545), ('co2 emission', 0.554)], [('msw fuel', 0.5784), ('efficiency
gasification', 0.5352), ('gasification efficiency', 0.532), ('fuel power', 0.5146), ('gasification modeling', 0.5073)],
[('waste transportation', 0.5937), ('waste density', 0.5807), ('facility waste', 0.577), ('waste sector', 0.5726),
('waste disposal', 0.5696)], [('degradation indonesia', 0.5317), ('emissions indonesia', 0.5061), ('worlds greenhouse',
0.4723), ('environment science', 0.4523), ('environment laws', 0.4505)], [('energy sources', 0.4908), ('energy
resources', 0.46), ('sources energy', 0.4581), ('setting indonesia', 0.4363), ('world energy', 0.424)], [('european
forestry', 0.5821), ('forestry sector', 0.5427), ('bioeconomy forest', 0.5361), ('imports forest', 0.4961),
('transformation bioeconomy', 0.4951)], [('pollution crop', 0.6563), ('crop residues', 0.5883), ('climate emissions',
0.5845), ('gases pollutants', 0.5798), ('residues emission', 0.5741)], [('indonesia implementers', 0.5116), ('quality
infrastructure', 0.4995), ('infrastructure documentation', 0.4876), ('smes industry', 0.4085), ('machine industry',
0.3999)], [('surabaya city', 0.4783), ('city indonesia', 0.4643), ('households surabaya', 0.4544), ('economy
environment', 0.3991), ('impact surabaya', 0.3913)], [('climate parameters', 0.5441), ('weather parameters', 0.5301),
('climate factors', 0.4911), ('weather climate', 0.4714), ('profile atmosphere', 0.4503)], [('emissions economies',
0.6728), ('economies emission', 0.6249), ('emissions trade', 0.6002), ('emissions perspective', 0.5809), ('emissions
capita', 0.5764)], [('availability fuels', 0.5609), ('fishermen fuel', 0.5553), ('boats income', 0.5137), ('cost fuels',
0.507), ('fuels fact', 0.485)], [('forest emission', 0.5968), ('area peat', 0.5458), ('forest psf', 0.529), ('swamp
forest', 0.5183), ('contributor greenhouse', 0.5161)], [('emissions globally', 0.6841), ('emissions implementation',
0.6383), ('carbon emissions', 0.6305), ('emissions standards', 0.6227), ('emissions emissions', 0.6076)], [('emissions
wastewater', 0.635), ('emissions waste', 0.6145), ('emissions facilities', 0.5973), ('emission waste', 0.5848),
('wastewater management', 0.5777)], [('greenhouse gases', 0.5543), ('emissions model', 0.5437), ('globally emissions',
0.5255), ('climate health', 0.523), ('climate aerosols', 0.5215)], [('emissions maintenance', 0.5603), ('pavement
maintenance', 0.5376), ('maintenance rehabilitation', 0.487), ('pavement research', 0.4818), ('rehabilitation pavement',
0.4792)], [('rice nitrogen', 0.7025), ('rice environmentally', 0.6482), ('rice conservation', 0.6447), ('environmentally
rice', 0.6327), ('rice production', 0.6266)], [('production indonesia', 0.5988), ('indonesia processes', 0.5285),
('indonesias sago', 0.5284), ('starch indonesia', 0.5033), ('sago production', 0.4926)], [('infrastructure forests',
0.6212), ('threats forests', 0.6021), ('deforestation degradation', 0.5708), ('regulations forests', 0.5594),
('degradation forest', 0.5583)], [('fish wastes', 0.6127), ('landfills indonesias', 0.5739), ('wastes landfills',
0.5343), ('wastes landfill', 0.5265), ('generation wastes', 0.5255)], [('biomass waste', 0.5894), ('emissions process',
0.555), ('emissions impact', 0.5549), ('emissions life', 0.5419), ('process emissions', 0.5396)], [('sugar emissions',
0.643), ('sugar consumption', 0.5733), ('fuels sugar', 0.5711), ('sugar production', 0.5562), ('sugar cultivation',
0.5303)], [('wastewaters feasibility', 0.6521), ('strategy wastewater', 0.6498), ('wastewater pollution', 0.6469),
('wastewater management', 0.6352), ('techniques wastewater', 0.6282)], [('cpo production', 0.4986), ('biomass
production', 0.4932), ('pome biogas', 0.4862), ('biogas production', 0.4637), ('biogas biomass', 0.4575)], [('analysis
crosscountry', 0.5691), ('crosscountry analysis', 0.5369), ('countries data', 0.52), ('coal production', 0.5021),
('support coal', 0.496)], [('diet study', 0.5767), ('balance diets', 0.5741), ('diets food', 0.5599), ('studi diet',
0.5448), ('food consumption', 0.543)], [('sectors externalities', 0.5273), ('externalities model', 0.5108), ('variables
greenhouse', 0.4872), ('externalities degradation', 0.4686), ('sector degradation', 0.4562)], [('wood biomass', 0.7372),
('biomass utilization', 0.6543), ('energy wood', 0.6515), ('wood source', 0.5738), ('biomass', 0.5693)], [('reputation
facts', 0.5336), ('companies sensitivity', 0.5227), ('emissions disclosure', 0.5074), ('emission disclosure', 0.4984),
('ghg emissions', 0.4805)], [('committees greenhouse', 0.4648), ('role women', 0.4358), ('risk committees', 0.4183),
('committee company', 0.4128), ('directors gender', 0.406)], [('pollutants countries', 0.5827), ('resources pollutants',
0.5817), ('resources emissions', 0.5299), ('countries emissions', 0.4924), ('resource pressures', 0.4743)],
[('profitability rating', 0.5802), ('ghg emissions', 0.5275), ('companies emission', 0.5248), ('company rating', 0.5226),
('performance profitability', 0.517)], [('road construction', 0.5485), ('emission impact', 0.5096), ('construction
research', 0.4873), ('ghg emission', 0.4791), ('construction role', 0.4712)], [('ghg emissions', 0.5939), ('emissions
industry', 0.5567), ('emissions impact', 0.5109), ('emissions rti', 0.4822), ('emissions', 0.462)], [('emissions waste',
0.5762), ('waste management', 0.5577), ('indonesia emissions', 0.5324), ('emissions treatment', 0.5252), ('waste
treatment', 0.4967)], [('rubber production', 0.7478), ('rubber manufacturing', 0.7459), ('demand rubber', 0.734),
('production rubber', 0.7304), ('rubber future', 0.7284)], [('efficiency timber', 0.6553), ('timber utilization',
0.6152), ('impacts logs', 0.6047), ('timber average', 0.5771), ('timber harvesting', 0.5646)], [('emissions reductions',
0.5839), ('reductions emissions', 0.5637), ('actions emissions', 0.5491), ('emissions deforestation', 0.5463),
('commitments emissions', 0.5367)], [('emissions forests', 0.633), ('area forests', 0.5614), ('timber harvesting',
0.5546), ('forests timber', 0.5323), ('area timber', 0.507)], [('option drilling', 0.4589), ('risk drilling', 0.454),
('drilling project', 0.4452), ('drilling handling', 0.4387), ('drilling campaign', 0.4261)], [('emissions deforestation',
0.5937), ('carbon market', 0.522), ('degradation indonesia', 0.513), ('environmentally need', 0.5111), ('environmentally
behaviours', 0.5103)], [('emissions logistics', 0.5825), ('emissions framework', 0.5276), ('emissions companies',
0.5067), ('companies indonesia', 0.4955), ('emissions transport', 0.4874)], [('emissions vehicles', 0.4869), ('road
data', 0.4735), ('vehicles study', 0.4704), ('road condition', 0.4606), ('sample road', 0.4579)], [('ash composition',
0.5716), ('concrete ash', 0.5438), ('ash strength', 0.5392), ('ash basedgeopolymer', 0.5194), ('cement production',
0.4823)], [('committees greenhouse', 0.4657), ('role women', 0.439), ('risk committees', 0.4195), ('committee company',
0.4099), ('directors gender', 0.4072)], [('infrastructure indonesia', 0.6122), ('indonesia road', 0.5454), ('road
infrastructure', 0.5331), ('pavement construction', 0.5248), ('pavement projects', 0.5226)], [('country deforestation',
0.6235), ('deforestation rate', 0.6065), ('rate deforestation', 0.6005), ('indonesias economy', 0.5991), ('deforestation
analysis', 0.5979)], [('consequence deforestation', 0.6551), ('forest degradation', 0.6257), ('deforestation study',
0.6254), ('deforestation forest', 0.6086), ('deforestation', 0.594)], [('fvwmanagement landfill', 0.5682), ('waste fvw',
0.5595), ('waste landfill', 0.5567), ('landfill digestion', 0.542), ('vegetable waste', 0.5406)], [('vehicles fuels',
0.5376), ('pollution warming', 0.5121), ('vehicles pollution', 0.5076), ('fuels contributors', 0.4806), ('vehicles
electric', 0.4759)]]

Cluster 0:
[[('biodiesel burden', 0.6814), ('biodiesel production', 0.6777), ('impact biodiesel', 0.6614), ('oil biodiesel',
0.6274), ('biodiesel oil', 0.6145)], [('biomass indonesia', 0.6269), ('palm biomass', 0.6023), ('sustainability biomass',
0.6004), ('source biomass', 0.587), ('biomass potential', 0.5638)], [('indonesias palm', 0.6245), ('indonesia palm',
0.6222), ('palm nation', 0.5789), ('palm oil', 0.5611), ('decades malaysia', 0.5404)], [('biogas energy', 0.5373), ('ghg
emissions', 0.5372), ('emissions benefits', 0.5342), ('biogas power', 0.5281), ('aspects biogas', 0.5068)], [('indonesia
sustainability', 0.6573), ('sustainability indonesia', 0.6517), ('sustainability policy', 0.5582), ('palm plantations',
0.5562), ('impacts sustainability', 0.55)], [('palm expansion', 0.567), ('palm plantations', 0.5473), ('palm plantation',
0.529), ('oil palm', 0.5188), ('palm land', 0.4789)], [('costs deforestation', 0.6256), ('emissions deforestation',
0.6142), ('deforestation scenario', 0.5608), ('deforestation option', 0.5571), ('deforestation forest', 0.508)],
[('impact greenhouse', 0.5376), ('emissions indonesia', 0.5162), ('peatlands carbon', 0.5032), ('ecosystem impact',
0.4953), ('policies conservation', 0.4706)], [('emissions indonesia', 0.5681), ('indonesia cpo', 0.5543), ('indonesia
issues', 0.5105), ('indonesia variables', 0.4967), ('palm oil', 0.4839)], [('impact production', 0.5276), ('emissions
indonesia', 0.5181), ('product impact', 0.5082), ('indonesia emission', 0.4989), ('industry impact', 0.4926)],
[('indonesia palm', 0.6179), ('malaysia palm', 0.6117), ('palm country', 0.5793), ('industry indonesia', 0.5594), ('palm
oil', 0.5582)], [('demands sustainability', 0.5286), ('sustainability consumers', 0.5045), ('sustainability study',
0.486), ('indonesia consumer', 0.4773), ('products indonesia', 0.4692)], [('palm peatlands', 0.6361), ('peatlands
environmentally', 0.5619), ('palm cultivation', 0.5382), ('cultivation peatlands', 0.5225), ('indonesias oil', 0.5023)],
[('biodiesel palm', 0.5847), ('palm oil', 0.5091), ('palm plantation', 0.4825), ('effect biodiesel', 0.4807), ('oil
indonesia', 0.4607)], [('biodiesel indonesia', 0.6348), ('government biodiesel', 0.6034), ('biodiesel plan', 0.5687),
('factors biodiesel', 0.5473), ('biodiesel adoption', 0.5358)], [('biofuel role', 0.646), ('issues biodiesel', 0.6187),
('biodiesel supply', 0.5995), ('biodiesel industry', 0.5939), ('sources biodiesel', 0.5839)], [('palm oil', 0.5148),
('oil palm', 0.5068), ('certification palm', 0.5011), ('products palm', 0.4675), ('rbd palm', 0.4292)], [('palm oil',
0.5907), ('oil indonesia', 0.5403), ('sustainability palm', 0.5157), ('production palm', 0.5148), ('indonesia palm',
0.4816)], [('palm oil', 0.5011), ('palm moratorium', 0.501), ('regulation palm', 0.4949), ('management palm', 0.4761),
('indonesia palm', 0.4742)], [('palm plantations', 0.5463), ('emissions plantations', 0.5457), ('palm production',
0.5223), ('palm cultivation', 0.5019), ('emissions fertilizers', 0.4956)], [('crops biodiesel', 0.506), ('impacts
biodiesel', 0.5033), ('biodiesel consumption', 0.4608), ('consumption crops', 0.4567), ('change crops', 0.4555)],
[('production palm', 0.55), ('palm oil', 0.5497), ('products indonesia', 0.5406), ('oil palm', 0.5356), ('palm products',
0.5284)], [('footprint biodiesel', 0.6796), ('cycle biodiesel', 0.6419), ('biodiesel stage', 0.6344), ('biodiesel
production', 0.619), ('production biodiesel', 0.6105)], [('agroforestry models', 0.505), ('land monoculture', 0.4644),
('plantations monoculture', 0.4548), ('palm plantations', 0.4427), ('ecosystem services', 0.4285)], [('sector biodiesel',
0.6479), ('biodiesel industry', 0.631), ('industry biodiesel', 0.6226), ('impacts biodiesel', 0.6158), ('order
biodiesel', 0.6081)], [('emissions land', 0.5459), ('footprint indonesian', 0.5073), ('ghg emissions', 0.4912),
('indonesia ghg', 0.4893), ('greenhouse gas', 0.4779)], [('palm oil', 0.5409), ('emissions palm', 0.519), ('oil supply',
0.4861), ('production oil', 0.4837), ('sources palm', 0.4643)], [('palm expansion', 0.5756), ('palm cultivation',
0.5267), ('palm plantation', 0.5141), ('palm crops', 0.511), ('oil palm', 0.4842)], [('palm plantations', 0.5538),
('palms ecosystem', 0.5192), ('oil palms', 0.5044), ('plantation oil', 0.4948), ('soil greenhouse', 0.4942)],
[('emissions contribution', 0.6262), ('gas emissions', 0.6186), ('emissions petroleum', 0.6172), ('emissions oil',
0.6121), ('emissions importance', 0.61)]]

Gambar 2.2. Hasil frasa yang dapat dibentuk (2)

Adapun setiap elemennya, terdapat sebuah tupel yang berisi frasa dan nilai desimal, pada elemen kedua atau nilai desimal tersebut menunjukkan tingkat kepentingan atau relevansi frasa tersebut dalam teks dalam rentang nilainya 0 sampai 1. Semakin tinggi nilai desimal, semakin penting dan relevan frasa tersebut dalam konteks teks yang sedang dianalisis.