

# ANALYSIS OF RELEVANT KEYWORDS TO GREENHOUSE GAS FACTORS IN INDONESIA USING TEXT MINING AND LUHN'S THEORY

## THESIS PROJECT

### Abstract

The causes of Greenhouse Gas (GHG) emissions have become a pressing global issue to address, including in Indonesia. This study aims to identify the factors contributing to GHG emissions in Indonesia through the analysis of abstract data from literature studies of the past five years. The approach involves techniques such as text mining, clustering, and selection using Luhn's theory. By employing the K-Means algorithm in clustering, abstract data is categorized into several clusters based on content similarity. The Luhn's theory-based selection method also identifies the most relevant keywords that represent the relevance of GHG emissions in Indonesia. The analysis results indicate that factors related to the aspects relevant to GHG emissions in Indonesia include deforestation, biodiesel, CO<sub>2</sub>/carbon, and electricity. This study provides a comprehensive understanding of the factors associated with the GHG aspect in Indonesia, offering essential insights for policymakers, experts, and readers to develop effective mitigation strategies.

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

### Objective

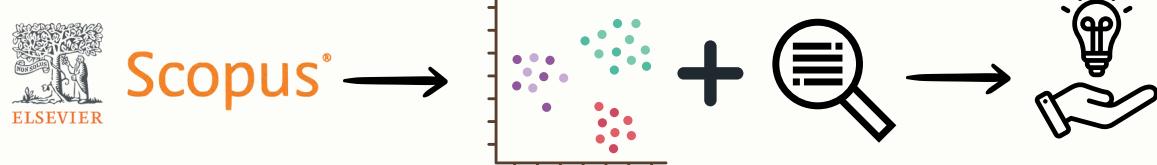


This research is expected to provide significant benefits in determining the factors contributing to Greenhouse Gas (GHG) emissions in Indonesia based on the latest information from the Scopus database, a prominent source of scholarly literature.

By employing a combined approach of quantitative and qualitative methods using text mining with NLP, along with K-Means clustering and the Luhn theory, this study is anticipated to streamline and enhance the identification of relevant keywords pertaining to GHG emissions. This will ensure a valid and fact-based understanding of the subject matter. Moreover, the outcomes of this research can serve as a foundation for developing improved policies and strategies, driven by the relevance of findings, to address climate change challenges and safeguard environmental sustainability for a better future, particularly concerning GHG emissions in Indonesia.

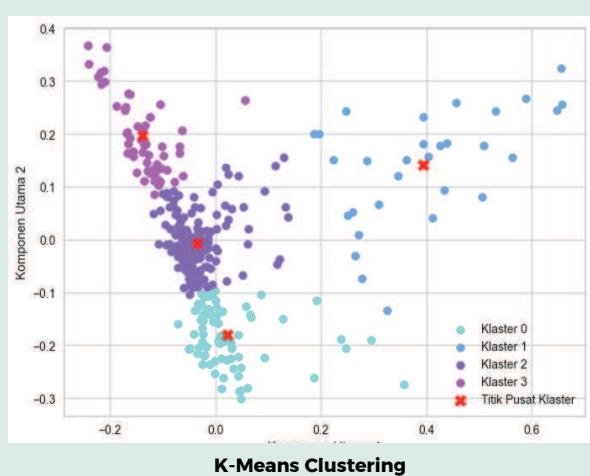
### Method

This research follows a structured workflow comprising several stages. The initial step involves identifying issues related to Greenhouse Gas (GHG) emissions in Indonesia. Subsequently, a data collection phase ensues, involving the compilation of abstract data from research journals using the Scopus platform. Following this, the data undergoes text mining techniques to extract relevant keywords.

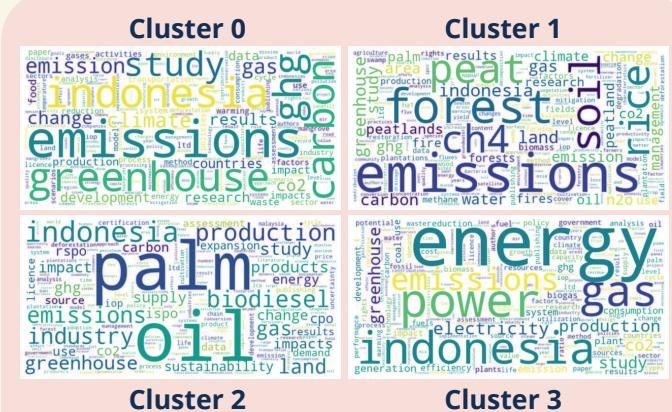


Next, the fundamental keywords generated through text mining, the selection of keywords is performed by assessing their relevance to the causal factors of GHG emissions. This selection is informed by analytical methods such as TF-IDF, PCA, K-Means, and Luhn theory. The final phase entails drawing conclusions from the research findings.

### Result and Discussion



The clustering algorithm designed to group data into similar clusters based on specific features, which represent the importance of words in each document relative to the entire dataset



Visual Representative with Word Cloud  
Represents the frequency of each cluster's keywords before processing to be selected with Luhn's Theory also qualitative analysis to generate GHG's factors and subfactors

Faktor dan 10 Subfaktor Aspek GRK di Indonesia berdasarkan Kata Kunci yang Dihasilkan		
Klaster	Faktor	Subfaktor
Cluster 0	Deforestation	emissions peat, deforestation degradation, mangrove deforestation, mangrove carbon, deforestation mangroves, emissions peatlands, vegetation indices, land fires, emissions peat, forestry land
Cluster 1	Biodiesel	biodiesel production, biomass indonesia, indonesian palm, palm peatlands, biodiesel palm, palm oil, impacts biodiesel, footprint biodiesel, palm expansion, biofuel role
Cluster 2	CO <sub>2</sub> / Carbon	buildings co2emissions, co2 cement, co2resistant cementing, coal power, waste composition, co2 atmosphere, dependence fuels, contributors carbon, waste business, pollutants countries
Cluster 3	Electricity	electricity consumption, policies energy, buildings energy, boiler efficiency, electricity supply, energy policy, generator efficiency, indonesian electricity, energy sources, capacity biogas

#### Factors and 10 Subfactors of GHG Aspects in Indonesia based on Generated Keywords

After selecting keywords from the combined results of clustering techniques and the principles of the Luhn theory, based on sources from literature studies, the aspects and factors contributing to GHG in Indonesia, these results can effectively serve as a foundation to assess the current situation by utilizing the last 5 years of literature that addresses the subject or urgency of GHG in Indonesia.

### Conclusions

- Identified Key Factors: Analysis revealed significant keywords contributing to Greenhouse Gas (GHG) causes in Indonesia, grouped into four clusters: deforestation, biodiesel, CO<sub>2</sub>/carbon, and electricity.
- Insights for Action: Valuable insights gained provide a basis for addressing and mitigating GHG impacts, focusing on the four identified clusters.
- Effective Techniques: Text mining, clustering, and Luhn's theory effectively pinpointed essential factors and sub-factors in Indonesian GHG emissions.
- Addressing Challenges: Targeted approaches are essential for deforestation, biodiesel promotion, CO<sub>2</sub> reduction, and improved electricity strategies to manage GHG emissions.
- Informed Solutions: The clustering of relevant keywords enhances understanding of complex GHG challenges, guiding informed policy-making and sustainable solutions.

### Suggestions

- Expand data sources for a comprehensive understanding.
- Collaborate with experts for validation and credibility.
- Consider temporal, geographical, and climatic aspects for a holistic impact analysis.
- Develop effective mitigation strategies based on identified factors.
- Formulate sustainable policies for enduring solutions.
- Stay updated with evolving methodologies and technologies.
- Contribute to public awareness and education initiatives.
- Embrace interdisciplinary collaborations for diverse insights.
- Engage stakeholders for successful strategy implementation.
- Encourage continuous research to support sustainable solutions.