

PROSPECTS AND CHALLENGES OF ALTERNATIVE TRANSPORTATION IN INDONESIA: UNDERSTANDING BASED ON NEWS EXTRACTION ABOUT TECHNOLOGY AND ACCEPTANCE OF ELECTRIC VEHICLES

EVITA PURNANINGRUM*, ABDURAKHMAN & NANANG SUSYANTO

ABSTRACT

The global transition to electric vehicles (EVs) presents both opportunities and challenges for emerging markets like Indonesia. This study applies text mining techniques—TextRank and Latent Dirichlet Allocation (LDA)—to analyze news articles and uncover key themes shaping EV adoption in Indonesia. TextRank analysis reveals increasing public awareness of EVs' environmental benefits but highlights affordability concerns and knowledge gaps as major barriers to widespread adoption. LDA results emphasize the central role of battery technology, Indonesia's nickel resources, and government policies in driving the industry forward. However, regulatory uncertainties, cost competitiveness, and environmental risks associated with nickel mining and battery production pose significant challenges. These findings provide a data-driven perspective on Indonesia's EV landscape, offering insights that can inform policymakers, industry stakeholders, and researchers in advancing sustainable transportation solutions.

Keywords: Electric Vehicles; Latent Dirichlet Allocation; TextRank; News; Preferences; Adoption

1. Introduction

Electric vehicles (EVs) represent a reemergence of technology with the potential to significantly impact the transportation sector. Instead, the evolution of electric vehicle technology is not a completely new technology in the transportation sector; rather, it combines modern technology with environmental precautions. Widespread adoption of EVs could significantly reduce transportation emissions, which are directly linked to local air pollution and contribute to global warming, and promote the deployment of renewable energy. Moreover, EVs are expected to gain popularity due to their economic and environmental benefits, as evidenced by the increasing consumer interest and government initiatives in several countries (Mandys 2021). Moreover, several countries have switched to electric vehicle systems, emphasizing cost reductions, infrastructure development, consumer information, and communications, and environmental impact activities. Therefore, adopting electric vehicle technology demands a wide range of inputs from several sources, making the process complex and requiring a multifaceted approach. Challenges include the development of a robust charging infrastructure, overcoming limitations in battery range, and ensuring cost-effective production of EVs.

Effective and efficient policymaking for EVs industry requires a comprehensive understanding of factors influencing customer adoption. Over the last few decades, a vast body of research exists on this topic, highlighting the complexity of the decision-making process. A thorough review of the literature will help to integrate the findings and gain a deeper understanding of the subject (Wang *et al.* 2018; Zhao *et al.* 2022). Previous research showed that customer adoption of electric vehicles is a complex process driven by a variety of factors. Furthermore, recent research suggests that preventing customer churn – the phenomenon where consumers abandon EVs after purchase – is crucial for long-term success. Optimizing policies to address factors like charging infrastructure limitations and ongoing vehicle costs could be instrumental in keeping consumers engaged with EV technology.

Previous research on EV adoption have relied on traditional survey-based methods, which can be time-consuming, expensive, and limited in their ability to capture the dynamic nature

of public opinion. To overcome these limitations, this study employs text mining techniques to analyze a vast corpus of online news articles related to EVs in Indonesia. Specifically, we utilize TextRank (Ao *et al.* 2020; Jain *et al.* 2024; Liu *et al.* 2024), a graph-based algorithm, to extract key sentences and summarize news articles, providing a concise overview of the dominant narratives surrounding EVs. The TextRank model leverages Word Embedding kind of Global Vectors for Word Representation (GloVe) (Rakshit & Sarkar 2025) to assess the significance of sentences within the news summaries. Additionally, we leverage Latent Dirichlet Allocation (LDA), a topic modeling technique, to identify and explore the prominent themes and discussions within the extracted news data. By adjusting LDA hyperparameters (Barna & Knap 2023), we aim to optimize topic coherence and gain deeper insights into the factors shaping EV adoption in Indonesia.

This study contributes to the growing body of literature on EV adoption by offering a data-driven understanding of the Indonesian context. The insights generated from this research can inform strategic decision-making for both public and private stakeholders, facilitating the development of effective policies and business strategies that promote sustainable transportation solutions in Indonesia.

2. Methodology

2.1. Data

The data for this study were collected through web scraping of online news sources using the Google News and Newspaper3k modules in Python on June 8, 2024. The scraping process was conducted on June 8, 2024, focusing on extracting relevant information, including publication date, media outlet, headline, and full text of news articles from Google News. A two-tiered keyword strategy was implemented to ensure comprehensive data retrieval. The first tier employed the broad keyword phrase "Electric Vehicle Industry" (denoted as Y), providing a wide net for relevant articles. Recognizing the multifaceted nature of public discourse surrounding electric vehicles, the second tier utilized three specific keywords: "Acceptance" ($X1$), "Public Perception" ($X2$), and "Transportation Electrification" ($X3$). These keywords were chosen based on their relevance to key themes within the broader topic of EV adoption. "Acceptance" and "Public Perception" are frequently used in social science and marketing research to gauge public attitudes and willingness to adopt new technologies (Zieffle *et al.* 2014; Boudet 2019; Lashram & Alkabaa 2024). "Transportation Electrification" is a more technical term often used in policy and industry discussions, representing a broader system-level shift toward electric transportation (Clairand & González-Rodríguez 2022). These keywords allow us to capture a broader range of perspectives, from individual consumer attitudes to broader societal and infrastructural considerations.

Given the limited availability of articles related to the electric vehicle (EV) industry in Indonesia, no restrictions were placed on the publication date. This decision allowed the inclusion of all relevant articles, regardless of their publication date, maximizing the dataset's scope. As a result, 260 articles were retrieved for the general keyword (Y), and 86, 80, and 86 articles were obtained for the specific keywords $X1$, $X2$, and $X3$, respectively.

During the initial web scraping process, raw data included articles with incomplete or irrelevant content, as well as duplicates. Rigorous data cleaning and pre-processing steps were applied to address this, including the removal of articles with missing information (e.g., absent headlines or full text) and those irrelevant to the research focus (e.g., articles unrelated to the Indonesian context or electric vehicle industry). These steps inevitably resulted in data loss, as some articles did not meet the inclusion criteria. Table 1 summarizes the raw data collected, the number of articles retained after pre-processing (referred to as 'Retained Data'), and the percentage of articles discarded during pre-processing due to data quality issues ('Data Loss').

Table 1: Comparison dataset from GoogleNews

Keywords	Raw Data Collected	Retained Data	Data Loss
EV Industry (Y)	260	235	9.61%
Acceptance ($X1$)	86	76	11.63%
Public Perception ($X1$)	80	71	11.25%
Transportation Electrification ($X3$)	86	78	9.30%

2.2. News extraction

The proliferation of online news articles provides abundant information but also presents challenges in navigating large amounts of text. Text Summarization techniques (El-Kassas *et al.* 2021), such as TextRank (Mallick *et al.* 2019; Zhang *et al.* 2020), address this issue by extracting key sentences to produce concise summaries (Moratanch & Chitrakala 2017). TextRank is a graph-based ranking algorithm that represents text as a network and adapted from Google’s PageRank algorithm (Mishra 2024), widely used in natural language processing (NLP) tasks. It constructs a network where sentences are represented as nodes, and edges between nodes indicate semantic similarity or relevance.

The importance of each sentence (node) is calculated iteratively using Eq. 1, which determines the centrality score $W(X_i)$ for a node X_i :

$$W(X_i) = (1 - d) + d * \sum_{j \in In(X_i)} \frac{W(X_j)}{|Out(X_j)|} \quad (1)$$

This equation reflects the idea that a sentence’s importance depends on the importance of other sentences linking to it. The centrality score of a node, denoted as $W(X_i)$, which represents the importance of sentence X_i within the text network. This score is determined by considering the importance of sentences that “link” to X_i , reflecting the principle that a sentence’s significance depends on the importance of the sentences that reference or relate to it. Specifically, $In(X_i)$ represents the set of sentences that have directed edges pointing towards X_i —these are the sentences that “influence” X_i . $Out(X_j)$, on the other hand, represents the set of sentences that X_j points to—those that X_j “influences.” The size of this set, denoted as $Out(X_j)$, is used for normalization. The damping factor, d (typically between 0.85 and 1), plays a crucial role in the algorithm. It represents the probability of continuing to “follow” links within the graph, preventing the algorithm from getting trapped in cycles and balancing local and global importance. A higher d gives more weight to the influence of connected nodes. The graph structure employed in this study is represented as $G = (X, Y)$, where X denotes the set of nodes, each corresponding to a sentence or textual unit, and Y represents the directed edges that capture the connections between nodes. Edge weights signify the similarity or relevance between sentences, reflecting their semantic relationships.

Figure 1 visually illustrates this structure, with nodes (X_1 to X_7) representing individual sentences and directed edges indicating their relationships. Edge labels, such as $in(X)$ and $out(X)$, specify incoming and outgoing connections, respectively. For example, $in(X_3)$ denotes sentences (e.g., X_1 and X_4) that reference or are semantically linked to X_3 , while $out(X_1)$ shows that X_1 influences the context or meaning of another sentence, such as X_3 . This graph representation enables the TextRank algorithm to compute sentence importance based on connectivity and relevance, effectively identifying the most central sentences for text summarization and ensuring the summary accurately reflects the key points of the original document.

Furthermore, extracting latent themes from textual data can be instrumental in aiding decision-makers with critical information. In this study, a modified version of Latent Dirichlet Allocation

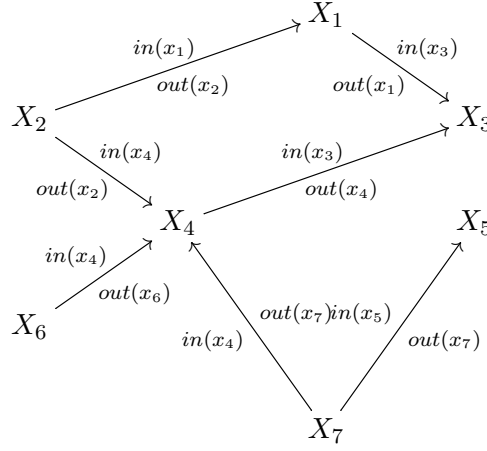


Figure 1: Example of a directed graph used in TextRank

tion (LDA), as proposed by Řehůřek and Sojka (2010), is employed to identify themes. LDA, a Bayesian topic modeling technique, operates by modeling the probabilistic relationships between words and topics using a multinomial likelihood distribution and a Dirichlet prior distribution. The parameters α and η play critical roles in regulating the model's behavior: α controls the distribution of topics across documents, while η governs the distribution of words within topics.

Eqs. 2 and 3 define the Dirichlet priors for topic-word and document-topic distributions:

$$\beta_k \sim \text{Dirich}(\eta) \quad \text{for } k = 1, 2, \dots, K \quad (2)$$

$$\theta_d \sim \text{Dirich}(\alpha) \quad \text{for } d = 1, 2, \dots, D \quad (3)$$

Here, β_k is a vector representing the probability of each word being in topic k , and θ_d is a vector representing the probability of each topic being in document d . η and α are the hyperparameters that influence these distributions. Eq. 4 provides the joint probability distribution of the LDA model, linking the observed data (words w) with the latent variables (topics z , topic distributions θ , and word distributions β):

$$p(w, z, \theta, \beta | \alpha, \eta) = \prod_{k=1}^K p(\beta_k | \eta) \prod_{d=1}^D p(\theta_d | \alpha) \prod_{n=1}^{N_d} p(z_{d,n} | \theta_d) p(w_{d,n} | z_{d,n}) \quad (4)$$

This distribution is composed of several factors: $p(\beta_k | \eta)$, the probability of the word distribution for topic k given hyperparameter η ; $p(\theta_d | \alpha)$, the probability of the topic distribution for document d given hyperparameter α ; $p(z_{d,n} | \theta_d)$, the probability of assigning topic $z_{d,n}$ to the n -th word in document d given θ_d ; and $p(w_{d,n} | z_{d,n})$, the probability of observing word $w_{d,n}$ given its assigned topic $z_{d,n}$. These components together define the likelihood of observing the words in the corpus given the latent topic structure.

The posterior distribution, as expressed in Eq. 5, computes the probabilities of the latent variables given the observed data:

$$p(z, \theta, \beta) = \frac{p(z, \theta, \beta, w)}{p(w)} \quad (5)$$

This equation is derived from Bayes' theorem and expresses the probability of the latent variables (topics z , topic distributions θ , and word distributions β) given the observed words w . Calculating this posterior directly is computationally intractable. Therefore, as discussed by Řehůřek (2011) and implemented by (Xue *et al.* 2024), we use variational inference with Jensen's inequality to approximate the posterior and derive the estimates $p(\text{word}|\text{topic})$, the probability of each word belonging to a specific topic, and $p(\text{topic}|\text{document})$, the probability of each topic appearing in a document. These outputs are directly derived from the distributions defined in Equations 2, 3, and 4.

Table 2 provides a matrix representation of these results. The 'Topic Distribution over Documents' illustrates how documents are probabilistically associated with topics, while the 'Word Distribution over Topics' shows how topics are characterized by their constituent words. For example, $p(\text{word}|\text{topic})$ reflects the contribution of individual words to a topic (as modeled by β), and $p(\text{topic}|\text{document})$ indicates the relevance of topics to specific documents (as modeled by θ). These relationships ensure that the table aligns with the parameters and formulas described above, illustrating how LDA systematically connects documents, topics, and words.

Table 2: Topic distribution over documents and word distribution over topics

Documents	Topics				
	z_1	z_2	z_3	z_4	z_5
d_1	$\beta_{1,1}$	$\beta_{1,2}$	$\beta_{1,3}$	$\beta_{1,4}$	$\beta_{1,5}$
d_2	$\beta_{2,1}$	$\beta_{2,2}$	$\beta_{2,3}$	$\beta_{2,4}$	$\beta_{2,5}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
d_D	$\beta_{D,1}$	$\beta_{D,2}$	$\beta_{D,3}$	$\beta_{D,4}$	$\beta_{D,5}$
Words					
	$\theta_{1,1}$	$\theta_{1,2}$	$\theta_{1,3}$	$\theta_{1,4}$	$\theta_{1,5}$
w_1	$\theta_{1,1}$	$\theta_{1,2}$	$\theta_{1,3}$	$\theta_{1,4}$	$\theta_{1,5}$
w_2	$\theta_{2,1}$	$\theta_{2,2}$	$\theta_{2,3}$	$\theta_{2,4}$	$\theta_{2,5}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
w_V	$\theta_{V,1}$	$\theta_{V,2}$	$\theta_{V,3}$	$\theta_{V,4}$	$\theta_{V,5}$

By combining TextRank and LDA, we aim to extract the most crucial information and uncover the underlying themes within the news articles related to electric vehicles (EVs) in Indonesia. TextRank, a graph-based algorithm, helps us identify the most crucial sentences within each article by considering their connections and importance in the overall text. The TextRank model leverages GloVe (Global Vectors for Word Representation), a pre-trained word representation tool from the Stanford NLP Group, to represent words as vectors that capture their semantic relationships, enabling a more nuanced assessment of each sentence's significance. On the other hand, LDA (Latent Dirichlet Allocation), a topic modeling technique, allows us to discover hidden topics or themes within the collection of news articles. It achieves this by grouping words that frequently occur together, assuming that these words represent a common theme. The LDA model's behavior is guided by hyperparameters, which we adjust to ensure meaningful and coherent topic discovery. Furthermore, we apply sentiment analysis using N-gram and VADER (Valence Aware Dictionary and sEntiment Reasoner) (Hutto & Gilbert 2014; Borg & Boldt 2020), a lexicon and rule-based sentiment analysis tool available in the NLTK library, to the news headlines. This helps us gauge the overall sentiment (positive, negative, or neutral) associated with electric vehicles in the Indonesian news landscape and identify any potential biases or misleading headlines.

The overall research process, as illustrated in Figure 2, consists of two main stages. In the first stage (Figure 2a), TextRank is employed to extract and summarize relevant news articles based on specific keywords related to electric vehicles in Indonesia. The sentences within each article are weighted using GloVe word embeddings to prioritize the most informative ones. In the second stage (Figure 2b), LDA is applied to the collection of weighted sentences to uncover the latent thematic structures present in the Indonesian EV industry news coverage.

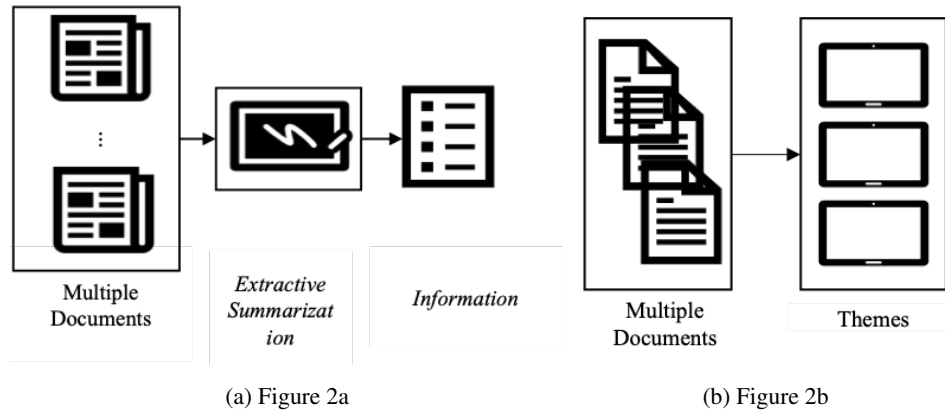


Figure 2: Two-stage text analysis approach

3. Results and Discussion

3.1. Sentiment analysis across headlines, full articles, and summaries

This subsection presents a comparative analysis of sentiment results derived from news headlines, full articles, and summaries. All three data types—headlines, article content, and summaries—were extracted simultaneously from Google News for each news item. Each set of data is inherently connected, as they are sourced from the same articles. The inclusion of these three elements allows for a deeper understanding of how sentiment varies across different textual levels of the news and ensures the analysis captures variations in tone and coherence. Additionally, this approach enables us to investigate whether headlines act as “clickbait,” a growing concern given that many readers only engage with headlines rather than full articles.

Sentiment analysis was conducted using VADER, a robust tool within the NLTK library that leverages a sentiment dictionary to calculate sentiment scores. VADER assigns scores by summing the weighted frequencies of words or phrases containing sentiment across each text unit. Table 3 presents the sentiment distribution—categorized as negative, neutral, and positive—across headlines (“Title”), full articles (“A”), and summaries (“S”) for various keywords related to electric vehicles (EVs) in Indonesian news. The “Change” column highlights the percentage point difference in sentiment between headlines and their corresponding full articles or summaries.

The analysis revealed a minimal change in negative sentiment between headlines and full articles (-3.90%) or summaries (-0.83%), suggesting that “clickbait” headlines are not prevalent in EV-related news in Indonesia. On the other hand, positive sentiment showed a significant increase when moving from headlines to full articles (average increase: 62.96%) and summaries (average increase: 56.44%). For instance, for the keyword “EV Industry,” positive sentiment rose from 38.30% in headlines to 94.04% in full articles and 88.09% in summaries. This trend underscores the tendency of headlines to present a less optimistic tone, while the full articles and summaries deliver more positive sentiment about EV-related topics.

Table 3: Comparative sentiment analysis of headlines and articles

Keywords	Sentiment Class	Headlines (%)	A or S (%)	Change (%)
EV Industry	Negative	4.26	A 5.11	↑ 0.85
			S 8.09	↑ 3.83
	Neutral	57.45	A 0.85	↓ 56.6
			S 3.83	↓ 53.62
	Positive	38.30	A 94.04	↑ 55.74
			S 88.09	↑ 49.79
Acceptance	Negative	9.21	A 6.58	↓ 2.63
			S 10.53	↑ 1.32
	Neutral	63.16	A 0.00	↓ 63.16
			S 5.26	↓ 57.89
	Positive	27.63	A 93.42	↑ 65.79
			S 84.21	↑ 56.58
Public Perception	Negative	12.68	A 1.41	↓ 11.27
			S 4.23	↓ 8.45
	Neutral	54.93	A 0.00	↓ 54.93
			S 4.23	↓ 50.70
	Positive	32.39	A 98.59	↑ 66.20
			S 91.55	↑ 59.15
Transportation Electrification	Negative	5.13	A 2.56	↓ 2.56
			S 5.13	0
	Neutral	61.54	A 0.00	↓ 61.54
			S 1.28	↓ 60.26
	Positive	33.33	A 97.44	↑ 64.10
			S 93.59	↑ 60.26

*A = full articles, S = article summaries

Moreover, the analysis showed a decrease in neutral sentiment as the text length increased, indicating that headlines often adopt a neutral tone, whereas full articles and summaries delve into detailed information, expressing stronger sentiment. These findings demonstrate VADER's capability to capture sentiment differences across text formats and lengths effectively. By analyzing the sentiment in Indonesian news media, this study offers indirect evidence of growing public support and optimism toward electric vehicles.

3.2. Descriptive statistics

A descriptive analysis of news data retrieved from Google News was conducted using keywords associated with Electric Vehicle (EV) adoption in Indonesia. N-gram analysis, specifically focusing on Bi-grams (two-word sequences) and Tri-grams (three-word sequences), was employed to identify frequently co-occurring word sequences within specific and general EV news categories. This analysis helps to uncover key phrases and collocations that shed light on the prominent themes and discussions surrounding EVs in the Indonesian context. The choice to focus on Bi-grams and Tri-grams is motivated by their ability to capture meaningful phrases and contextual relationships between words, providing more nuanced insights compared to single word analysis. The key findings from the N-gram analysis are visualized in Figure 3.

Notably, the prominent N-grams suggest a focus on the entire EV supply chain, encompassing elements ranging from raw materials to finished vehicles, with a particular emphasis on the availability of critical components such as batteries. Additionally, the analysis reveals the

potential environmental benefits of EVs, including reduced carbon emissions and contributions to sustainability goals. Figure 3 provides a visual representation of the most frequent Bi-grams and Tri-grams identified in the news articles, offering a snapshot of the key themes and concerns prevalent in the Indonesian EV discourse.

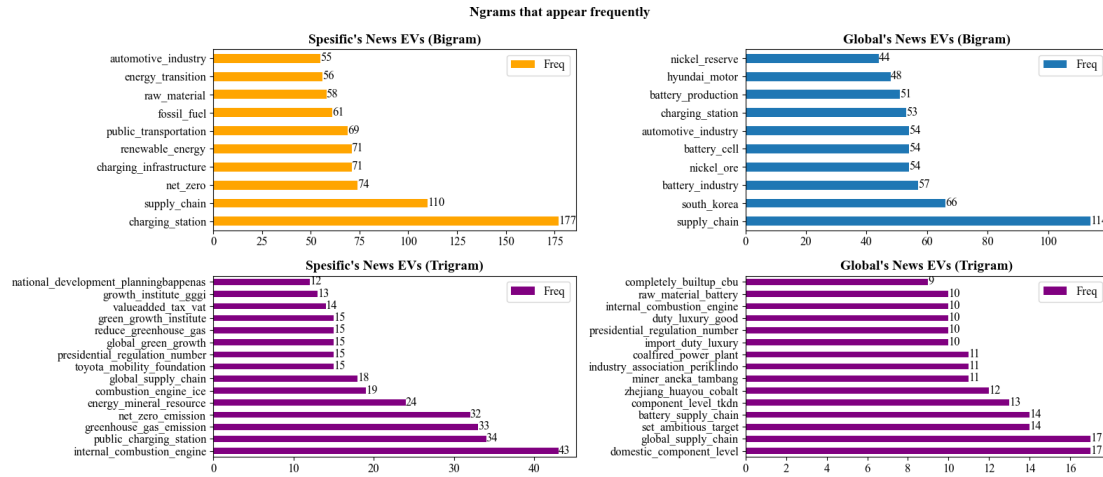


Figure 3: Key themes in Indonesian EV news: an N-gram analysis

3.3. TextRank analysis

The TextRank algorithm, as described in the methodology, constructs a graph where nodes represent sentences and edges capture their relationships based on semantic similarity. The PageRank algorithm (Eq. 1) is then applied to compute TextRank scores, which indicate the relative importance of each sentence within the document. Higher scores suggest sentences that are more central and interconnected, while lower scores indicate less connected or peripheral sentences. The average TextRank scores for each section (headline, summary, article) and subject (Acceptance (X1), Public Perception (X2), Transportation Electrification (X3)) are presented in Table 4.

Table 4: Average TextRank scores by section and subject

Subject	Headline	Summary	Article
Acceptance (X1)	0.0130	0.0028	0.0004
Public Perception (X2)	0.0135	0.0029	0.0004
Transportation Electrification (X3)	0.0128	0.0027	0.0003

The TextRank scores indicate the degree of inter-sentence connectivity within each section and subject. The relatively low average scores (ranging from 0.0003 to 0.0135) suggest weak semantic linkages between sentences, potentially reflecting fragmented discourse or a lack of coherence in the analyzed documents. The decreasing trend from headlines to summaries to full articles suggests that while titles tend to be more information-dense and interconnected, the main body of the text may contain more loosely connected statements.

In the context of EV acceptance, a low level of inter-sentence connectivity might imply that discussions on this topic are scattered, lacking a clear, cohesive narrative. This could reflect inconsistent messaging, diverse stakeholder opinions, or fragmented information in public discourse, potentially hindering a unified understanding and acceptance of EV technology.

3.4. LDA analysis

We employed an iterative approach to determine the optimal number of topics K for our LDA model. This involved testing different topic numbers (from 1 to 20) and selecting the configuration that yielded the highest coherence score (c_{max}), which measures how semantically related the words within each topic are. Table 5 summarizes the results for different combinations of hyperparameters α and η , which influence topic distribution. The hyperparameters α and η regulate how topics are distributed across documents and how words are assigned to topics:

- **Symmetric:** Assigns equal probability to all topics (for α) or words (for η). This assumes a balanced topic-word distribution across documents.
- **Asymmetric:** Allows topic distributions to vary across documents, meaning some topics dominate more than others.
- **Auto:** Learns α and η from the data, adapting the distribution dynamically.

Table 5: Descriptive statistics for hyperparameter

Alpha (α)	Eta (η)	modeK	min c_{max}	c_{max}	max c_{max}		s
					value	K	
symmetric	symmetric	7	0.3392	0.3677	0.3931	7	0.0133
	auto	3	0.3484	0.3752	0.4185	3	0.0217
asymmetric	symmetric	3	0.3350	0.3833	0.4751	3	0.0316
	auto	3	0.3392	0.3706	0.4445	3	0.0235
auto	symmetric	2	0.3486	0.3735	0.4331	3	0.0242
	auto	2	0.3515	0.3770	0.4297	3	0.0215

In our study, the highest average coherence (0.3833) and a high maximum coherence (0.4751) were achieved with α set to "asymmetric" and η set to "symmetric," with $K = 3$. This suggests that three topics best represent the dominant themes in the data. The choice of "asymmetric" for α implies that we allowed the model to learn potential biases in the prevalence of different EV-related themes in the news articles. For example, perhaps news coverage disproportionately focuses on technological aspects of EVs more than policy considerations. The "symmetric" setting for η suggests that, within each topic, we initially assumed all EV-related terms were equally likely until the model learned otherwise. These choices, while data-driven, are not directly measuring "acceptance" of EVs. Instead, they allow us to discover what aspects of EVs are being discussed and how those discussions are structured.

Table 6 illustrates the probabilistic relationship between words and the three topics identified by the LDA model. The table presents the top 10 most probable words for each topic, offering insights into the thematic focus of each. It's important to understand that these are probabilities, not absolute assignments. A word can belong to multiple topics, but with varying degrees of probability. The listed words are those with the highest probability of association with each respective topic.

The results presented in this section, derived from our TextRank and LDA analyses, provide a data-driven foundation for understanding the prospects and challenges facing the Indonesian EV industry. The key themes and insights identified through these computational methods inform the discussion in the following section, where we explore the multifaceted landscape of EV adoption in Indonesia and consider the implications of our findings for policy and industry.

4. Prospects and Challenges of Alternative Transportation in Indonesia

Building upon the insights from our TextRank and LDA analyses, this section contextualizes the key findings within the broader landscape of Indonesia's EV industry. The TextRank anal-

Table 6: Identified topics and their top words

Topic	Top Words
Topic 1 (Policy and Market Dynamics)	incentive, industry, automotive, market, tax, regulation, investment, vin-fast, minister, domestic
Topic 2 (EV Technology and Manufacturing)	battery, service, meet_firm, auto, nickel, maker, market, neta, time, price
Topic 3 (Resource Extraction and Sustainability)	nickel, battery, mining, mineral, plant, investment, energy, produce, ore, industry

ysis, which highlights societal acceptance through keywords such as 'acceptance' and 'public perception,' identifies the most salient concerns and opportunities shaping EV adoption. These insights are explored in Section 4.1. Meanwhile, the LDA topic modeling uncovers recurring themes, including government incentives, battery production, and environmental considerations, which are further analyzed in Section 4.2 to highlight key industry trends and challenges. Finally, Section 4.3 discusses the broader implications of these findings and their alignment with existing research.

4.1. Challenges and opportunities in Indonesia's EV adoption

The transition to electric vehicles (EVs) in Indonesia presents both opportunities and challenges, with public perception playing a critical role in shaping adoption. One of the most significant concerns highlighted in the analysis is the high upfront cost of EVs compared to gasoline vehicles. This aligns with global trends, where affordability remains a major barrier—only 4% of consumers currently own an EV, while 12% are considering a purchase. Recognizing this challenge, the Indonesian government has introduced measures such as VAT reductions to make EVs more accessible. However, public perception analysis suggests that financial incentives alone may not be sufficient; raising awareness about long-term cost benefits, including lower maintenance and fuel savings, is equally important.

Infrastructure development is another crucial factor influencing EV adoption. The government's push to electrify public transportation, particularly buses, is a step in the right direction. However, successful implementation depends on securing large-scale fleet orders to achieve meaningful impact. Furthermore, the expansion of charging station networks is essential to mitigate range anxiety, one of the key deterrents to EV adoption. A well-developed infrastructure will not only enhance convenience for users but also signal strong government commitment to supporting the EV ecosystem.

4.2. Key themes in Indonesia's EV industry

Topic modeling using LDA reveals several dominant themes in Indonesia's EV discourse. Across all model configurations, battery technology and nickel resources emerge as central topics, reflecting Indonesia's strategic role as a leading global supplier of nickel—a key component in EV battery production. Terms such as battery, nickel, plant, factory, and mining consistently appear across different hyperparameter settings, with document frequencies ranging from 54 to 233. This underscores the nation's ambition to position itself as a key player in the EV supply chain, leveraging its natural resources to attract investment and establish itself as a regional hub for EV manufacturing.

Government policies, including investment incentives and market development initiatives, also appear prominently in the topic distribution, with document counts ranging from 181 to 211. These topics reflect ongoing efforts to accelerate industry growth through regulatory support and financial incentives. However, challenges persist. Discussions surrounding regulation,

tax, and price (emerging in models with $\alpha=\text{auto}$, $\eta=\text{symmetric}$ and $\alpha=\text{symmetric}$, $\eta=\text{auto}$, with document counts between 29 and 96) indicate that cost competitiveness and policy uncertainty remain significant concerns. Additionally, the appearance of terms such as environmental and pollution (notably in the $\alpha=\text{symmetric}$, $\eta=\text{symmetric}$ model with a document count of 48) suggests that sustainability issues—particularly regarding nickel mining and battery production—are an emerging focal point in public discourse.

4.3. Broader implications and research alignment

These findings align with previous research on EV adoption, reinforcing the complexity of consumer decision-making in this sector. The identification of affordability as a major barrier is consistent with global studies emphasizing cost concerns as a key factor influencing EV uptake. Likewise, the importance of a well-developed charging infrastructure mirrors existing research highlighting its role in reducing range anxiety and improving user confidence.

Additionally, the study contributes to ongoing discussions about policy optimization for EV adoption. By highlighting the significance of economic incentives, regulatory frameworks, and sustainability considerations, the results provide valuable insights for policymakers aiming to balance industrial growth with environmental responsibility. Addressing challenges such as infrastructure gaps, tax policies, and public awareness campaigns will be crucial in ensuring long-term engagement with EV technology and preventing consumer disengagement.

In summary, this study not only confirms known challenges in the EV sector but also expands on them by incorporating insights from media analysis and topic modeling. By integrating public discourse with data-driven findings, it offers a more nuanced understanding of the factors shaping Indonesia's transition to electric mobility.

5. Conclusion

This study provides a comprehensive analysis of Indonesia's evolving electric vehicle (EV) industry by leveraging TextRank and Latent Dirichlet Allocation (LDA). The TextRank analysis highlights a growing public awareness of EVs' environmental benefits but also underscores persistent concerns regarding affordability and knowledge gaps. Addressing these challenges requires a multifaceted approach, including consumer education initiatives, strategic infrastructure expansion, and policies that enhance EV affordability.

LDA analysis reveals dominant themes in Indonesia's EV discourse, emphasizing the critical role of battery technology, nickel resources, and government-backed market development. However, challenges related to regulatory frameworks, cost competitiveness, and environmental sustainability remain significant. Moving forward, a balanced strategy is essential—one that fosters technological innovation in battery production, ensures responsible nickel mining practices, and streamlines regulatory processes. By adopting a holistic approach, Indonesia can cultivate a thriving EV industry that aligns economic growth with environmental sustainability, paving the way for a cleaner and more resilient transportation future.

Acknowledgement

This work was supported by Universitas Gadjah Mada through the 2024 Final Project Recognition Program under Grant number 5286/UN1.P1/PT.01.03/2024.

References

- Ao X., Yu X., Liu D. & Tian H. 2020. News keywords extraction algorithm based on TextRank and classified TF-IDF. *2020 International Wireless Communications and Mobile Computing (IWCMC)*, pp. 1364–1369.

- Barna I. & Knap Á. 2023. Analysis of the thematic structure and discursive framing in articles about Trianon and the Holocaust in the online Hungarian Press using LDA topic modelling. *Nationalities Papers* **51**(3): 603–621.
- Borg A. & Boldt M. 2020. Using VADER sentiment and SVM for predicting customer response sentiment. *Expert Systems with Applications* **162**: 113746.
- Boudet H.S. 2019. Public perceptions of and responses to new energy technologies. *Nature Energy* **4**(6): 446–455.
- Clairand J.M. & González-Rodríguez M. 2022. What is the level of people's acceptance for electric taxis and buses? Exploring citizens' perceptions of transportation electrification to pay additional fees. *World Electric Vehicle Journal* **13**(1): 3.
- El-Kassas W.S., Salama C.R., Rafea A.A. & Mohamed H.K. 2021. Automatic text summarization: A comprehensive survey. *Expert Systems with Applications* **165**: 113679.
- Hutto C. & Gilbert E. 2014. VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media*, pp. 216–225.
- Jain R., Singh P. & Puri S. 2024. Summarization of daily news using TextRank and TF-IDF algorithm. *Proceedings of the 4th Congress on Intelligent Systems (CIS 2023)*, pp. 313–324.
- Lashram Y. & Alkabaa A.S. 2024. Navigating challenges in the transition to green transportation: A perception study exploring factors influencing drivers' intentions for electric vehicle adoption. *AIP Advances* **14**(3): 035009.
- Liu W., Sun Y., Yu B., Wang H., Peng Q., Hou M., Guo H., Wang H. & Liu C. 2024. Automatic text summarization method based on improved TextRank algorithm and k-means clustering. *Knowledge-Based Systems* **287**: 111447.
- Mallick C., Das A.K., Dutta M., Das A.K. & Sarkar A. 2019. Graph-based text summarization using modified TextRank. *Soft Computing in Data Analytics: Proceedings of International Conference on SCDA 2018*, pp. 137–146.
- Mandys F. 2021. Electric vehicles and consumer choices. *Renewable and Sustainable Energy Reviews* **142**: 110874.
- Mishra A. 2024. Data driven knowledge summarization of friction stir welded magnesium alloys literature by using natural language processing algorithms. *International Journal on Interactive Design and Manufacturing (IJDeM)* **18**(3): 1113–1119.
- Moratanch N. & Chitrakala S. 2017. A survey on extractive text summarization. *2017 International Conference on Computer, Communication and Signal Processing (ICCCSP)*, pp. 1–6.
- Rakshit P. & Sarkar A. 2025. A supervised deep learning-based sentiment analysis by the implementation of Word2Vec and GloVe embedding techniques. *Multimedia Tools and Applications* **84**(2): 979–1012.
- Řehůřek R. 2011. Scalability of semantic analysis in natural language processing. PhD Thesis. Masaryk University.
- Řehůřek R. & Sojka P. 2010. Software framework for topic modelling with large corpora. *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pp. 46–50.
- Wang N., Tang L. & Pan H. 2018. Analysis of public acceptance of electric vehicles: An empirical study in Shanghai. *Technological Forecasting and Social Change* **126**: 284–291.
- Xue Y., Kambhampati C., Cheng Y., Mishra N., Wulandhari N. & Deutz P. 2024. A LDA-based social media data mining framework for plastic circular economy. *International Journal of Computational Intelligence Systems* **17**(1): 8.
- Zhang M., Li X., Yue S. & Yang L. 2020. An empirical study of TextRank for keyword extraction. *IEEE Access* **8**: 178849–178858.
- Zhao X., Ma Y., Shao S. & Ma T. 2022. What determines consumers' acceptance of electric vehicles: A survey in Shanghai, China. *Energy Economics* **108**: 105805.
- Ziefle M., Beul-Leusmann S., Kasugai K. & Schwalm M. 2014. Public perception and acceptance of electric vehicles: Exploring users' perceived benefits and drawbacks. *Design, User Experience, and Usability. User Experience Design for Everyday Life Applications and Services (DUXU2014)*, pp. 628–639.

*Mathematics Department
Faculty of Mathematics and Natural Sciences
Universitas Gadjah Mada
Sekip Utara BLS 21
Yogyakarta, 55281
INDONESIA
E-mail: evitapurnaningrum0890@mail.ugm.ac.id*, rachmanstat@ugm.ac.id,
nanang_susyanto@ugm.ac.id*

Received: 18 October 2024
Accepted: 28 February 2025

*Corresponding author