Jurnal Manajemen Hutan Tropika, *32*(1), 63–72, January 2026
EISSN: 2089-2063
DOI: 10.7226/jtfm.32.1.63

Scientific Article
ISSN: 2087-0469

Media Framing of Human-*Orangutan* Interactions in Kalimantan: A Topic Modelling and Sentiment Analysis Approach

Tri Giyat Desantoro[1*], Lilik Budi Prasetyo[2], Muhammad Irfansyah Lubis[3]

[1]Tropical Biodiversity Conservation, Faculty of Forestry and Environment, IPB University, Academic Ring Road, Campus IPB Dramaga, Bogor, Indonesia 16680
[2]Department of Forest Resources Conservation and Ecotourism, Faculty of Forestry and Environment, IPB University, Academic Ring Road, Campus IPB Dramaga, Bogor, Indonesia 16680
[3]Research Centre for Biota Systems, National Research and Innovation Agency (BRIN), Cibinong, Indonesia 16911

Received April 28, 2025/Accepted August 27, 2025

Abstract

People's perceptions of orangutans strongly influence their responses in every interaction, which can have negative impacts on both humans and orangutans. Misperceptions, fear, or negative attitudes may increase the risk of conflict, injury, or retaliatory actions, thereby threatening conservation outcomes. Understanding community sentiment toward human–orangutan interactions is therefore essential for designing effective conservation strategies. However, direct data collection to understand community sentiment often faces cost and accuracy constraints. Therefore, an alternative approach that is more efficient and objective is needed to capture community sentiment toward human-orangutan interactions. This research offers latent topic modelling and sentiment analysis of online news articles as a novel method to understand these dynamics and support conservation efforts. A total of 11 latent topics were obtained from the news articles. Orangutan sightings and handovers of orangutans were the most frequently discussed in 2016. Palm oil plantations emerged as a prominent topic related to human-orangutan interaction incidents. Negative sentiment was predominantly linked to topic such as injured orangutans and orangutan sightings on plantations, whereas orangutan translocation received the highest positive sentiment score. This study highlights the potential of natural language processing for analyzing Indonesian language texts in conservation contexts, with applications extendable to broader environmental and forestry issues such as deforestation and wildlife hunting.

Keywords: *conservation response, bornean orangutan, topic modelling, text mining, media-based monitoring*

**Correspondence author, email*: *tgdesantoro@gmail.com*

Introduction

People's perspective on wildlife significantly influences their actions towards the wildlife they encounter. Wildlife may be perceived as harmful pests, animals that need to be protected, or dangerous threats (Suwannarong et al., 2023; Grundei et al., 2024). Negative perceptions can lead to harmful actions toward wildlife, most commonly found in large mammals, such as elephants, tigers, and *orangutans* (Fredriksson, 2005; Meijaard et al., 2011; Qomariah et al., 2018; Khatoon et al., 2022).

The bornean *orangutan* (*Pongo pygmaeus*) is one of the wildlife species that has the potential to interact with humans in the wild. These interactions can take many forms, such as captivity, hunting, fruit crops, or even just sightings of *orangutan*s near human settlements and activity areas (Campbell-Smith et al., 2010; Meijaard et al., 2011; Davis et al., 2013). Interactions that occur can be threatening for both parties. *Orangutan*s typically flee when they perceive a threat from humans, like other wildlife (Gaynor et al., 2025). However, it is undeniable that *orangutan*s can also attack humans when *orangutan*s feel pressured (Meijaard et al., 2011). Compared to *orangutans*, which choose to avoid humans, community sentiment influences the emotions and cognitions that are carried out during *orangutan* interactions (Aprillyasari, 2023). The human response can be hostile actions such as hunting or killing *orangutan*s perceived as agricultural pests, or seemingly compassionate acts such as capturing and keeping them as pets (Davis et al., 2013).

Efforts are needed to prevent negative actions that are detrimental to both parties. One crucial approach is to understand community sentiments. Different community sentiment made different acts to wildlife, *orangutans* especially (Kretser et al., 2009; Aprillyasari et al., 2025). Understanding community sentiment can help identify priority areas for targeted management and community engagement around orangutan habitats. Collecting sentiment-related data by visiting areas where negative human-*orangutan* interactions have occurred is relatively expensive and time-consuming. Mapping perception for the bornean *orangutan* case was conducted by visiting 725 villages and required 2 periods in different years (Abram et al., 2015). In addition, the community's explanation of the

Jurnal Manajemen Hutan Tropika, *32*(1), 63–72, January 2026
EISSN: 2089-2063
DOI: 10.7226/jtfm.32.1.63

Scientific Article
ISSN: 2087-0469

incident can have various versions because the length of time the event took place makes it difficult for respondents to remember.

Online news articles are an alternative documentation that records all events that occur on a regional, national, and global scale. Human-wildlife interaction is one topic that is not spared to be reported (Egri et al., 2021; Neo et al., 2023). However, news articles consist of various subtopics that can not represent the sentiment value of a single news story. News content contains latent topics, such as responses from the public, responses from authorised agencies, or even the opinions of the news writer. This creates a bias in determining the sentiment value that occurs in one piece of news content. News content analyses of human-predator conflicts have been executed with collaborative proofreading by the authors to ascertain sentiment and news framing (Ardiantiono et al., 2022). This research requires verification and alignment of perceptions between authors, but the initial subjectivity of the authors may influence the analysis process. Analyzing online news has been done for clustering news based on headline news about slow loris threats and conservation but not to reveal subtopics and assess news sentiment (Ihsan et al., 2024).

This research aims to analyze latent topics in news content of human-*orangutan* interactions and assess sentiment on these latent topics. this research initiates the development of natural language processing (NLP) for the Indonesian language for forestry, especially in human-*orangutan* interaction issues. Revealing latent topics and assessing latent topic sentiment in news articles makes it easier for parties, especially authorities, to understand incidents of negative human-*orangutan* interactions that affect *orangutan* conservation, thereby enriching knowledge for determining plans and directions for *orangutan* conservation policies in particular.

Methods

This research uses online news articles as the primary data sources, consisting in the form of text obtained from local and national news media in Indonesia that related to human-bornean *orangutan* interactions. The methodological process is illustrated in Figure 1.

Data collective News articles were collected by keyword using the Google search engine with the Boolean technique. Boolean techniques use logical operators such as AND, OR, and NOT to combine or exclude keywords, making search results more specific and relevant. We used a term combination, with 'AND' as the operation searching, namely 'orangutan' AND 'Kalimantan' AND 'Indonesia' AND 'year: 2011'. The news collected is in the Indonesian language regarding the interaction between humans and *orangutan*s from local to national news from June 2011 to June 2024 and stored in microsoft excel *file.csv* format. The study period was selected as it coincided with major conservation events, including the 2014 and 2019 forest fires (Hanifah et al., 2016; Wibowo et al., 2020) and the 2017 *orangutan* PHVA (Utami-Atmoko et al., 2017), which significantly influenced the dynamics of media coverage. Web page news was identified by its *html* address regarding news title, news publication

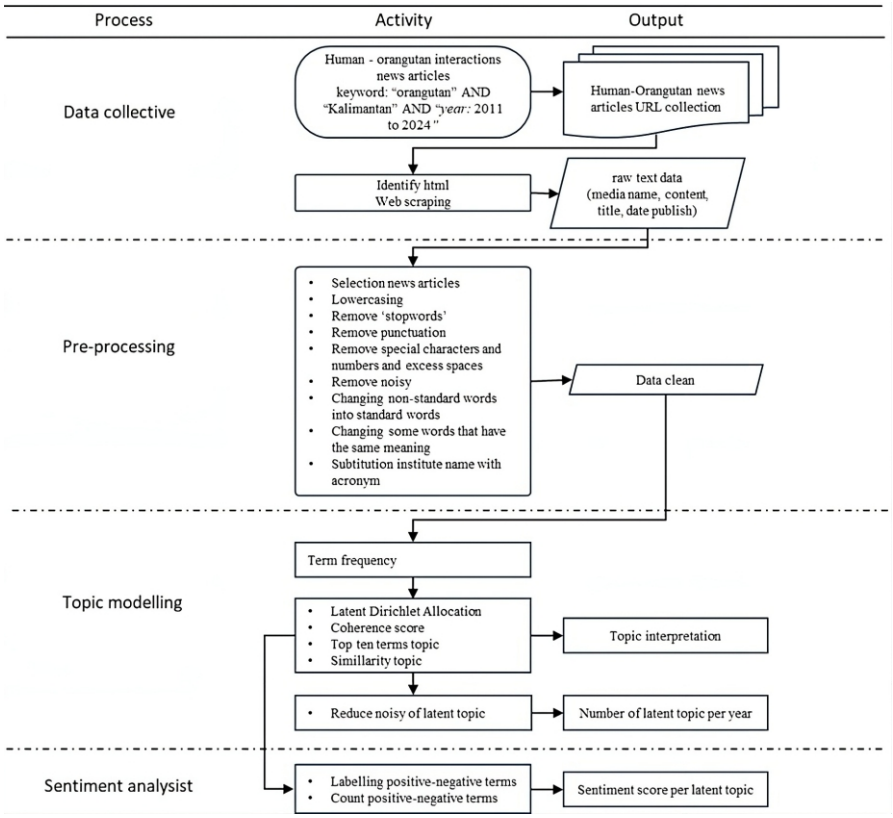

Figure 1 Research analysis process.

Jurnal Manajemen Hutan Tropika, *32*(1), 63–72, January 2026
EISSN: 2089-2063
DOI: 10.7226/jtfm.32.1.63

Scientific Article
ISSN: 2087-0469

date, and news content as a reference for the web scraping process or extractive text from web pages. Web scraping used Rstudio 4.3.2 and stored in micorsoft excel *file.csv*. The Rstudio library was used in this process, namely *dplyr* (Wickham et al., 2015), *polite* (Perepolkin, 2019), *rvest* (Wickham, 2014), *magrittr* (Bache & Wickham, 2022), and *stringr* (Wickham, 2009).

Pre-processing Published news articles often contain repetitive information due to similarities in reporting on the same incident, so it was necessary to filter out or eliminate duplicate articles. Selection of news articles based on the quantity of individual *orangutans*, sex of *orangutan*s, same location, and publication within 7 days would be eliminated, and only one news article would be chosen. Next, text processing on the news content is carried out, which aims to standardize the text.

(1) changing uppercase letters to lowercase letters on words (2) removing punctuation (3) removing connecting words such as 'di = in', 'yang = which', 'oleh = by', and 'karena = because' using stopwords (Tala, 2016); then (4) removing special characters, numbers, and excess spaces; and (5) removing noisy texts, noisy texts are very dangerous because they affect the expected results in the study. This noise is a text that is integrated with news content; it is necessary to check it manually. Noisy texts exhabit the pattern that can be detected in every news article (Yi et al., 2003). Some noisy texts were detected on news articles with word or sentence patterns such as 'read: ...'. "read more: ....". In addition, noise also appears in the form of advertisements that change into coding and language that is different from the main news content. (6) changing non-standard (*non-baku*) words into standard words (*baku*) or typo words such as 'menghimbau to mengimbau = appeal', 'aktifitas to aktivitas = activity', and 'cidera = cedera'. This technique aims to reduce the number of unique words (different words) that influence the next process. (7) changing some words with the same meaning, such as 'berangkat', 'menuju' to 'pergi' = 'go', 'mencatat', 'mencatatkan', 'mendokumentasikan' to 'mendokumentasi'. (8) substitute institution name and the name of the institution's affiliation with an acronym, as shown in Table 1.

Topic modelling Latent dirichlet allocation (LDA) tries to find latent topics in a collection of news articles. Each news article consists of several topics. Topics are composed of a collection of terms (Blei et al., 2003). The LDA process works by estimating the probability of terms (β) in news articles coming from certain topics. This research used collapsed gibbs sampling for LDA modelling with Rstudio 4.3.2 (Griffiths & Steyvers, 2004). Libraries used in Rstudio are *textmineR* (Jones, 2015), *topicmodels* (Grün & Hornik, 2011), *slam* (Hornik et al., 2024), and *ldatuning* (Nikita, 2015). The stages of work are to make modelling loops with the number of topics as many as 3 to 20 topics. The settings made in this study, namely seed (1234), which is a random number that is used as a starting point for making random numbers. It is intended that the process results do not change even though they are operated on other computers. The number of iterations or repetitions (2000) in the gibbs sampling model for more stable convergence; burn-in (500) to discard the number of initial results so as not to be biased; this is because the initial iteration has not yet obtained stable convergence, so 500 of 2,000 is not used; and thin (10) to take one sample from 10 repetitions so that the correlation is reduced between samples and inference results are improved.

The results of the LDA modelling process, which made 3 to 20 topics, are evaluated by coherence score. Coherence score shows how reasonable and meaningful the topics are based on the co-occurrence relationship of term probabilities in each topic (Mimno et al., 2011). Highest coherence score shows the better model. coherence score results are seen with a graph that shows the value increasing until it stagnates or decreases. Stagnant or decreased values indicate that terms are no longer meaningful because terms between topics have several terms that are the same in use. Furthermore, the terms from the selected number of topics are sorted and 10 terms with the largest probability are selected that have the most influence on the formation of each topic. This process was run using Rstudio with same the libraries with LDA modeling.

The selected number of topics was extracted from the 10 most influential terms and displayed by creating a graph. The graph was created with the library *ggplot2* (Wickham, 2016). Thereafter, topic similarity is measured by counting the number of terms in common from the top 10 unique terms between topics with jaccard coefficient. The result of the measurement is made into a heatmap. The greater the inter-topic value, the darker the color on the heatmap. Topic similarity means using the same words to build the topic, allowing two or more topics to merge. Heatmap was made with the library *pheatmap* (Kolde, 2019).

Table 1 Acronym of institution name

Insitute name	Acronym
Balai Konservasi Sumberdaya Alam; Seksi Konservasi Wilayah; Wildlife Rescue Unit	BKSDA
Borneo Orangutan Survival Foundation	BOSF
Centre Orangutan Protection	COP
Friends of National Park Foundation	FNPF
Orangutan Foundation Indonesia	OFI
Orangutan Foundation United Kingdom	OF UK
Sintang Orangutan Center	SOC
Yayasan International Animal Rescue; Yayasan Inisiasi Alam Rehabilitasi	YIARI
Yayasan Palung	YP

Jurnal Manajemen Hutan Tropika, *32*(1), 63–72, January 2026
EISSN: 2089-2063
DOI: 10.7226/jtfm.32.1.63

Scientific Article
ISSN: 2087-0469

The result of the analysis obtained low-to-high topic probabilities that appear in news articles. Low topic probabilities are eliminated to avoid confusion in interpretation and only present relevant topics in the article. Low probability elimination uses the default threshold value by calculating (1/K) with K = the number of topics selected. The results of latent topic occurrence for each article are presented in the form of a heatmap that shows the distribution of latent topic occurrence for each year. The darker the heatmap color, the higher the frequency of latent topic occurrence in articles each year. The *R* script used for the LDA process is available at the [link].

Sentiment analysis Sentiment analysis in this study relies on the Indonesian sentiment (Inset) lexicon dataset (Koto & Rahmaningtyas, 2017). The top 10 terms in each topic were labeled based on the inset lexicon. The process begins with identical word matches between the words of each topic and the inset lexicon. Unlabelled terms were then matched using the fuzzy matching method by calculating jaccard similarity using the stringdist library (van Der Loo, 2013). Next, the sentiment-labeled terms were manually curated to avoid misidentification in the labeling.

Sentiment measurement was calculated by comparing the proportion of positive words to negative words relative to the total number of words in each topic. Sentiment is positive if the proportion of positive values is greater than the proportion of negative values. Sentiment is negative if the proportion of positive values is less than the proportion of negative values and sentiment is neutral if the proportion of positive and negative values is balanced. Sentiment score shows the attitude, perception, and response to each latent topic.

Results

Human-*orangutan* interaction data There were 248 news articles obtained in the range of 2011 to 2024. The screening results to remove duplicated news articles (29%) and further processed were 175. Duplication of news articles that appear are usually news summaries of incidents that have occurred and are reported again every month. Typicallly, affiliated news media group also duplicate news articles. Figure 2 displays the distribution of deleted news articles.

The number of terms before text cleaning on news article content is 164,287 with a mean of 397 terms per news article. The distribution shows a minimum of 99 terms, a maximum of 1,846 terms, and a total of 14,551 unique terms. After preprocessing, the number of terms processed for topic modeling is 32,102 with a mean of 187 terms per topic, a minimum of 44 terms and a maximum of 756 terms with 4,478 unique terms. Noisy text data greatly affects the context of topic modeling in human-*orangutan* interaction content.

Topic model and selection Topic modeling produces term order that needs to be interpreted, especially in the 10 terms with the highest probabilistic value (Figure 3). Good topics are able to be interpreted according to word order. In addition, a coherence score values evaluate topic modeling by showing how words appear together in each topic. The coherence score results show that 11 topics get the highest score (0.073) compared to the number of other topics. The coherence score initially decreased from eight to ten topics and increased again at eleven. Thereafter, the coherence score stagnated and did not increase more than 0.073 (11 topics). Stagnant values indicate that the topic uses the same terms repeatedly, which makes interpretation difficult; additionally, having more than 11 topics is redundant due to the limited number of terms, resulting in topics that are only refined or narrowed down (Mimno et al., 2011).

Notable keyword Figure 4 displays the top 10 terms, along with their corresponding probability values, that constitute each topic. Word probability values range from 0.006 to 0.112, with an average probability value of 0.020. There are 82 unique terms of the 111 terms that make up the topic. 'Orangutan' is the term used for the formation of seven topics. In addition, the term 'hutan = forest' formed four topics, while the words 'warga = resident', 'bksda', 'individu', and 'kalimantan' formed three topics.

Similarity topic based jaccard similarity has shown that there is no dark color in the heatmap graphic (Figure 5). The Jaccard coefficient range is 0.05 to 0.25. Highest Jaccard coefficient between topic 2 and topic 3 (0.25). Both topics are related to *orangutan* sightings in community plantations and the existence oil palm plantation. This value shows the

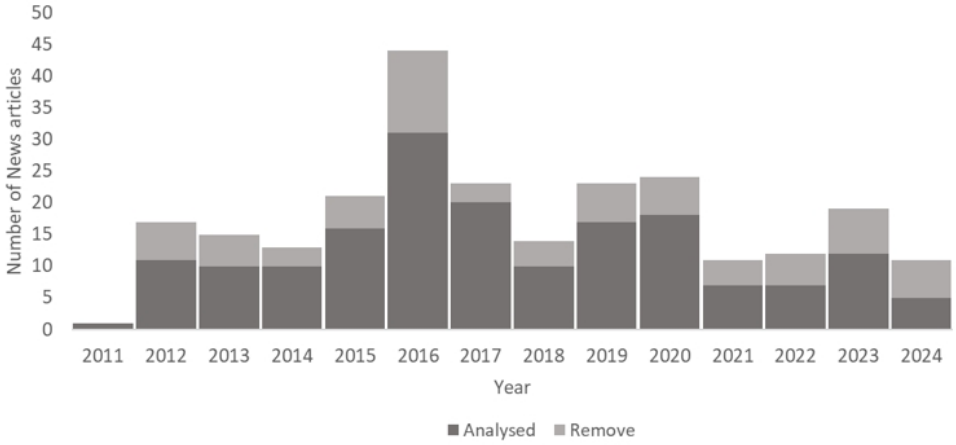

Figure 2 Number of news articles analyzed and remove per year.

Jurnal Manajemen Hutan Tropika, *32*(1), 63–72, January 2026
EISSN: 2089-2063
DOI: 10.7226/jtfm.32.1.63

Scientific Article
ISSN: 2087-0469

Figure 3 Coherence score per number of latent topics.

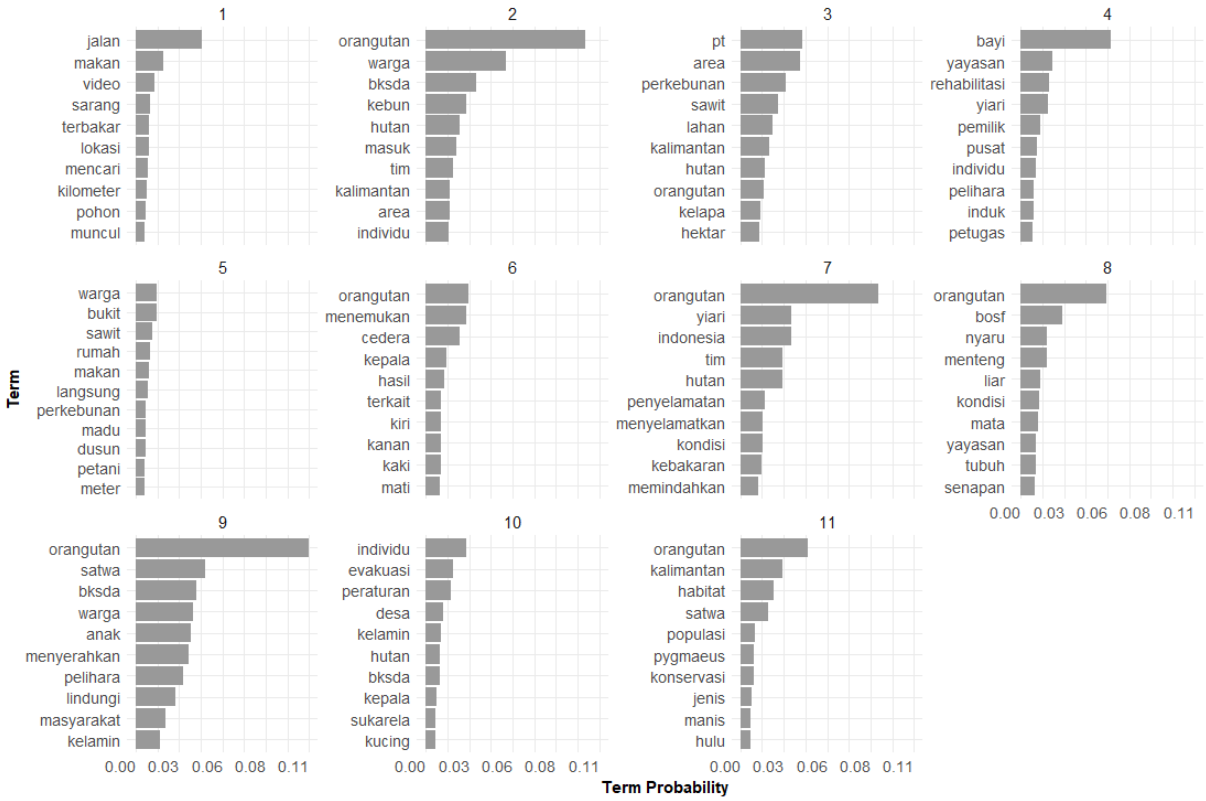

Figure 4 Term probability per topic.

potential for *orangutan* sightnings in community plantations, especially oil palm plantations. Low Jaccard coefficient (n < 0.5) indicates an interrelationship between topics or independence of these topics, so that these topics are not connected to each other or there is overfitting between topics.

Sentiment score The results of sentiment analysis of all terms in latent topics obtained 48 terms (23 negative terms and 25 positive terms) in Table 2. Negative terms consist of one to four terms per topic and positive terms consist of one to five terms. 'Hutan = forest' is a positive term that often appears in four latent topics, while the frequency of negative words is twice, namely 'kelamin = sex', "kepala = head', and 'tim = team'. The highest positive sentiment score is four, which is obtained on 'translocation orangutan' topic, while the negative sentiment is -2 on the *orangutan* injured condition topic.

Discussion

Topic modeling results cannot directly distinguish the opinions of the parties from the incidents of human-*orangutan* interactions in news content. However, the explicit interpretations of topic 10 and topic 11 are mitigation efforts in the form of appeals regarding regulations and

Jurnal Manajemen Hutan Tropika, *32*(1), 63–72, January 2026
EISSN: 2089-2063
DOI: 10.7226/jtfm.32.1.63

Scientific Article
ISSN: 2087-0469

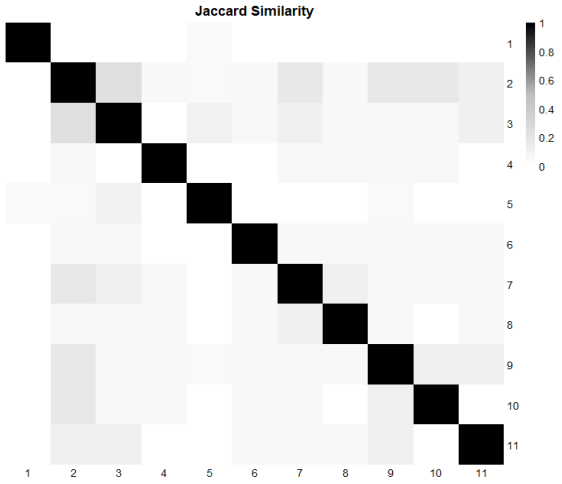

Figure 5 Jaccard similarity between topics.

Table 2 Sentiment score per topic

Topic name	Topic ID	Proportion sentiment terms	Sentiment score
Orangutan sighting in transportation area	1	0.60	0
Orangutan sighting in community plantation	2	0.50	-1
Oil palm plantations exist	3	0.30	1
Orangutan captive	4	0.70	1
Orangutan raiding crop	5	0.45	1
Orangutan hunting	6	0.80	0
Translocation of *orangutan*	7	0.60	4
Orangutan injured condition	8	0.40	-2
Handover *orangutan*	9	0.50	1
Wildlife protection regulation	10	0.60	0
Population and habitat of *orangutan*	11	0.40	0

education about *orangutan*s in news articles. The parties' opinions are more easily shown in the text process that comes from the comment column, such as e-commerce or tweets from 'X' social media, which are used to write down the opinions and impressions of its users compared to the parties' opinions in news content (Chamberlain, 2018; Ohtani, 2022).

Oil palm plantations are one of the latent topics identified in news articles on human-*orangutan* interactions. This reveals that landscape change is one of the factors driving human-*orangutan* interactions. These findings underscore that changes in land use and landscape structure, such as the expansion of roads and agricultural plantations, serve as major drivers of increased human-*orangutan* interactions. Human-*orangutan* conflict incidents are more likely to occur in human-modified environments, particularly in areas adjacent to roads and oil palm plantations (Oram et al., 2022; Risdiyanto et al., 2023). The fragmentation and reduction of natural habitats not only force *orangutan*s to range beyond forest boundaries but also elevate the chances of direct encounters with humans. Such landscape transformations disrupt *orangutan* movement patterns and resource availability, thereby intensifying competition and potential conflict. Thus, landscape-level planning and habitat connectivity are essential to mitigate these interactions and promote coexistence.

Topic interpretation Based on ten terms of each topic, the latent topics that emerged from topic modeling results can be interpreted and seen in Table 3. Most topics are closely related to the types of human-*orangutan* interactions, as seen in the topics on *orangutan* sightings in transportation areas, community gardens, and hunting activities. *Orangutan* sightings in areas of human acitivity were also reported by Campbell-Smith et al. (2011). Some topics also reflect landscape conditions, such as the presence of oil palm plantations, which can be a driving factor for these interactions. Number of human-*orangutan* interactions is linear with land cover change from forest to oil palm plantation (Ancrenaz et al., 2014; 2016). There are also topics that show the effect of interactions, like injured *orangutans* condition, and responses or treatments, like translocation and surrender. *Orangutan* translocation was still carried out in several incidents during the recent period (Figure 6). However, this approach is considered less effective as a long-term conservation strategy for *orangutan* populations (Sherman et al., 2020). All of these themes reflect the complexity of human-*orangutan* relationships that involve environmental factors and conflict, as well as mitigation and protection efforts.

Jurnal Manajemen Hutan Tropika, *32*(1), 63–72, January 2026
EISSN: 2089-2063
DOI: 10.7226/jtfm.32.1.63

Scientific Article
ISSN: 2087-0469

Sentiment scores On each latent topic, as shown in Table 3, showed a range of -2 to 4. There was a mismatch in sentiment scores between the concept of human-*orangutan* interaction and the sentiment scores obtained. For example, the topics "orangutan hunting" and "orangutan captive", which ethically and ecologically should be perceived negatively, have neutral and positive scores. This result indicates that, for the topic "orangutan captive," the number of positive terms exceeded the negative terms, whereas for "orangutan hunting," the number of positive and negative terms was equal. Some of the terms from the lexicon dataset by Koto and Rahmaningtyas (2017) are not appropriate for orangutan conservation. For example, *memelihara* = "captive" and *pemilik* = "owner" are categorized as positive terms. In contrast, the topic "translocation orangutans", a treatment and care effort that is often criticized as risky for animal welfare, actually has a high positive sentiment score (4) (Sherman et al., 2020; Oram et al., 2022). Topics such as "oil palm plantation existing" and "orangutan raiding crop" scored neutral and positive despite being potentially negative. The equal number of positive and negative terms for the category of oil palm plantation leads to a neutral value. Although the topic contains the term *sawit* (= oil palm), it is not included in the inset lexicon and is therefore considered not applicable. If seen as evidence of *orangutan* habitat being converted into plantation indicating the dominance of narratives of concern or human-*orangutan* conflict.

The inset lexicon by Koto and Rahmaningtyas (2017)

Table 3 Latent topic interpretation

Topic ID	Topic name	Themes
1	*Orangutan* sighting in transportation area	Interaction types
2	*Orangutan* sighting in community plantation	Interaction types
3	Oil palm plantation sexist	Landscape
4	*Orangutan* captive	Interaction types
5	*Orangutan* raiding crop	Interaction types
6	*Orangutan* hunting	Interaction types
7	Translocation of *orangutan*	Mitigation and handling
8	*Orangutan* injured condition	Interaction impact
9	Handover *orangutan*	Mitigation and handling
10	Wildlife protect ion regulation	Mitigation and handling
11	Population and habitat of *orangutan*	Mitigation and handling

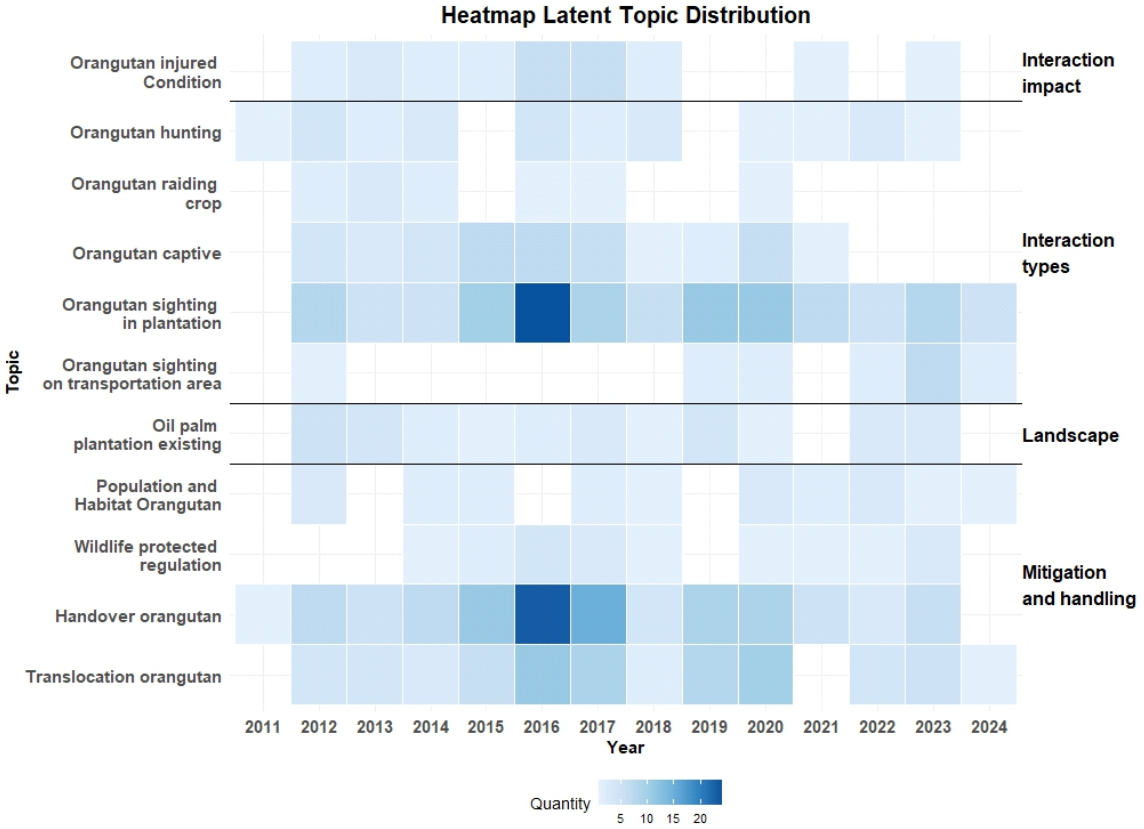

Figure 6 Heatmap latent topics per year.

serves as a general sentiment resource for the Indonesian language, but it lacks the ability to capture context-specific meanings, particularly in the domain of nature and wildlife conservation. A similar limitation for using lexicon was also reported by Khoo and Johnkhan (2017). The lexicon will be more accurate if it is developed from words that are appropriate for the context. The discrepancy in sentiment scores may be due to the fact that the lexicon inset used was not built on a more specific concept of human-wildlife interactions but rather a collection of texts derived from public opinion on social media.

Latent topic distribution Distribution of 11 temporal latent topics is shown by Figure 6. *Orangutan* sighting (n = 24/29%) and *orangutan* handover (n = 23/28%) are dominant topics in 2016. This graph shows that these two topics were the most highly reported in that year. *Orangutan* handover and *orangutan* sightings showed a more even occurrence throughout the year, indicating a persistent issue. *Orangutan* hunting is the lowest case yet still found in the last 14 years. Almost all topics emerged in 2016–2017. This coincided with the completion of the 2007–2017 *orangutan* conservation action plan strategy (SRAK) and publication of the *orangutan* population-habitat viability assessment (PHVA) report by the Ministry of Forestry at that time (Soehartono et al., 2007; Utami-Atmoko et al., 2017). Topics with the themes of interaction types and mitigation and handling are the ones that appear most frequently, reflecting the dynamics of direct interaction and ongoing conflict management efforts. *Orangutan* translocation remains a prefereed method for addressing negative interactions that may occur in areas where *orangutans* are present (Sherman et al., 2020).

Conservation implications The findings of this study demonstrate that topic and sentiment analysis of online media coverage can be used to monitor *orangutan*-related issues more rapidly and comprehensively. This information can assist stakeholders in identifying potential conflicts or threats to *orangutan*s before they escalate. The results also provide a foundation for developing more targeted conservation communication strategies. Thus, this research directly contributes to supporting adaptive and evidence-based *orangutan* conservation efforts.

Conclusion

This study discovered that online news articles can function as an alternative data source for recording incidents of negative interaction between human and *orangutan*, providing a more economical and expedient method in comparison to direct field-based data collection. Additionally, public perceptions and the evaluation of *orangutan* conservation policy implementation can be analyzed by juxtaposing trends in news coverage with pertinent policy advancements during the corresponding timeframe. This study underscores the subtantial limitations of latent topic modeling and sentiment analysis, despite their potential. Specifically, these methods fail to effectively distinguish between diverse stakeholder perspectives and factual descriptions of incidents. The shortcomings of sentiment analysis, in particular, arise from its reliance on a generic sentiment lexicon that inadequately reflects the complex nuances inherent in human–wildlife interactions. Future research should consider the development of domain-specific sentiment classifications tailored to conservation and human–wildlife conflict contexts. The creation of specialized sentiment frameworks for human–wildlife interactions presents a promising direction for enhancing the accuracy and ecological relevance of sentiment analysis in conservation research.

References

Abram, N. K., Meijaard, E., Wells, J. A., Ancrenaz, M., Pellier, A. S., Runting, R. K., Gaveau, D., Wich, S., Nardiyono, Tjiu, A., Nurcahyo, A., & Mengersen, K. (2015). Mapping perceptions of species' threats and population trends to inform conservation efforts: the Bornean *orangutan* case study. *Diversity and Distributions*, *21*(5), 487–499. https://doi.org/10.1111/ddi.12286

Ancrenaz, M., Gumal, M., Marshall, A. J., Meijaard, E., Wich, S. A., & Husson, S. (2016). *Pongo pygmaeus* (Errata version published in 2018). *The IUCN red list of threatened species* 2016: e.T17975A123809220. https://doi.org/10.2305/IUCN.UK.2016-1.RLTS.T17975A17966347.en

Ancrenaz, M., Oram, F., Laurentius, A., Lackman, I., Ahmad, E., Elahan, H., Kler, H., Abram, N. K., & Meijaard, E. (2014). Of pongo, palms and perceptions: A multidisciplinary assessment of bornean orang-utans *Pongo pygmaeus* in an oil palm context. *Oryx*, *49*(3), 465–472. https://doi.org/10.1017/S0030605313001270

Aprillyasari, W. M. (2023). Hubungan aktivitas antropogenik dengan interaksi orangutan dan manusia di Suaka Margasatwa Lamandau [thesis]. Bogor: IPB University.

Aprillyasari, W. M., Sunkar, A., & Rahman, D. A. (2025). Two sides of a coin: Do genders affect human-*orangutan* interactions in Lamandau Wildlife Reserve, Central Kalimantan? *Media Konservasi*, *30*(1), 73–83. https://doi.org/10.29244/medkon.30.1.73

Ardiantiono., Alfarisi, A. M., Ishaq, Y., Wijaya, R., Septian, R., Hadi, A. N., Surya, R. A., & Rahmi, T. (2022). Analisis konten pemberitaan interaksi manusia-satwa predator di Indonesia. *Jurnal Belantara*, *5*(2), 153–168. https://doi.org/10.29303/jbl.v5i2.874

Bache, S. M., & Wickham, H., (2022). *magrittr*: A forward-pipe operator for R. Retrieved from https://CRAN.R-project.org/package=magrittr

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. In T. G. Dietterich, S. Becker, & Z. Ghahramani (Eds.), *Advances in neural information processing systems 14: Proceedings of the 2001 conference* (pp. 601–608). MIT Press. https://doi.org/10.7551/mitpress/1120.003.0082

Campbell-Smith, G., Simanjorang, H. V. P., Leader-Williams, N., & Linkie, M. (2010). Local attitudes and perceptions toward crop-raiding by orangutans (*Pongo abelii*) and other nonhuman primates in northern Sumatra, Indonesia. *American Journal of Primatology*, *72*(10), 866–876. https://doi.org/10.1002/ajp.20822

Campbell-Smith, G., Campbell-Smith, M., Singleton, I., & Linkie, M. (2011). Raiders of the Lost Bark: Orangutan foraging strategies in a degraded landscape. *PLoS ONE*, *6*(6), Article e20962. https://doi.org/10.1371/journal.pone.0020962

Chamberlain, J. (2018). Using social media for biomonitoring: How Facebook, Twitter, Flickr and other social networking platforms can provide large-scale biodiversity data. *Advances in Ecological Research*, 59, 133–168. https://doi.org/10.1016/bs.aecr.2018.06.001

Davis, J. T., Mengersen, K., Abram, N. K., Ancrenaz, M., Wells, J. A., & Meijaard, E. (2013). It's not just conflict that motivates killing of orangutans. *PLoS ONE*, *8*(10), Article e75373. https://doi.org/10.1371/journal.pone.0075373

Egri, G., Han, X., Ma, Z., & Chakraborty, S. (2021). Classifying and analysing human-wildlife conflicts in India using news articles. *EasyChair* Preprint 5989. Retrieved from https://easychair.org/publications/preprint/8nS7

Fredriksson, G. (2005). Human-sun bear conflicts in East Kalimantan, Indonesian Borneo. *Ursus*, *16*(1), 130–137. https://doi.org/10.2192/1537-6176(2005)016[0130:HBCIEK]2.0.CO;2

Gaynor, K. M., Hayes, F. P., Manlove, K., Galloway, N., Benson, J. F., Cherry, M. J., Epps, C. W., Fletcher, J. r., Orrock, J., Smith, J. A., Aiello, C., Belant, J. L., Berger, J., Biel, M., Bright, J., Bump, J., Burchett, M., Butler, C., Carlson, J., ..., & Cross, C . P. (2025). The influence of human presence and footprint on animal space use in US national parks. *Proceedings of the Royal Society B*, *292*, Article 20251013. http://doi.org/10.1098/rspb.2025.1013

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(suppl_1), 5228–5235. https://doi.org/10.1073/pnas.0307752101

Grün, B., & Hornik, K. (2011). *topicmodels*: An package for fitting topic models. *Journal of Statistical Software*, *40*(13), 1–30. https://doi.org/10.18637/jss.v040.i13

Grundei, L. -L., Schöttes, F. M., Gethöffer, F., Tost, D., Kluge, L., Siebert, U., & Pees, M. (2024). Human–wildlife interaction–a social survey. *Animals*, *14*(5), Article 808. https://doi.org/10.3390/ani14050808

Hanifah, M., Syaufina, L., & Prasasti, I. (2016). The detection of burnt area using medium resolution satellite imagery of MODIS based on fire index approach. *Jurnal Pengelolaan Sumberdaya Alam dan Lingkungan*, *6*(1), 77–85. https://doi.org/10.29244/jpsl.6.1.77

Hornik, K., Meyer, D., & Buchta, C. (2024). *slam*: Sparse lightweight arrays and matrices. https://doi.org/10.32614/CRAN.package.slam

Ihsan, F., Suciadi, H., Latifatunnisa, H., Rosatika, F., Afandi, R., Stephen-Moore, R., & Sanchez, K. L. (2024, November 23–27). *Analysis of slow loris threats and conservation efforts from online news headlines using machine learning* [Conference presentation]. Asian Primates Symposium, Medan, Sumatra Utara, Indonesia.

Jones, T. (2015). *textmineR*: Functions for text mining and topic modeling. Version 3.0.5. https://doi.org/10.32614/CRAN.package.textmineR

Khatoon, R., Anwar, M., Nilon, C. H., & Gompper, M. E. (2022). Human-carnivore conflicts in a recently established Pakistani national park. *Human-Wildlife Interactions*, *16*(1), 144–157. https://doi.org/10.26077/4a09-b334

Khoo, C. S. G., & Johnkhan, S. B. (2017). Lexicon-based sentiment analysis: Comparative evaluation of six sentiment lexicons. *Journal of Information Science*, *44*(4), 491–511. https://doi.org/10.1177/0165551517703514

Kolde, R. (2019). *pheatmap*: Pretty heatmaps. Version 1.0.12. https://doi.org/10.32614/CRAN.package.pheatmap

Koto, F, & Rahmaningtyas, G. Y. (2017). Inset lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs. In *2017 International conference on Asian language processing* (pp. 391–394), Singapore. https://doi.org/10.1109/IALP.2017.8300625

Kretser, H. E., Curtis, P. D., Fracis, J. D., Pendall, R. J., & Knutuh, B. A. (2009). Factors affecting perceptions of human–wildlife interactions in residential areas of northern New York and implications for conservation. *Human Dimensions of Wildlife*, *14*(2), 102–118. https://doi.org/10.1080/10871200802695594

Meijaard, E., Buchori, D., Hadiprakarsa, Y., Utami-Atmoko, S. S., Nurcahyo, A., Tjiu, A., Prasetyo, D., Nardiyono, Christie, L., Ancrenaz, M., Abadi, F., Antoni, I. N. G., Armayadi, D., Dinato, A., Gumelar, E. P., Indrawan, T. P., Kussaritano, Munajat, C., Priyono, C. W. P., ..., & Mengersen, K. (2011). Quantifying killing of orangutans and human-orangutan conflict in Kalimantan, Indonesia. *PLoS ONE*, *6*(11), Article

e27491. https://doi.org/10.1371/journal.pone.0027491

Mimno, D., Wallach, H., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing semantic coherence in topic models. *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, *2011*, 262–272.

Neo, W. H. Y., Lubis, M. I., & Lee, J. S. H. (2023). Settlements and plantations are sites of human–tiger interactions in Riau, Indonesia. *Oryx*, *57*(4), 476–480. https://doi.org/10.1017/S0030605322000667

Nikita, M. (2015). *ldatuning*: Tuning of the latent dirichlet allocation models parameters. Version 1.0.2. https://doi.org/10.32614/CRAN.package.ldatuning

Ohtani, S. (2022). How is people's awareness of "biodiversity" measured? Using sentiment analysis and LDA topic modeling in the twitter discourse space from 2010 to 2020. *SN Computer Science*, *3*(5), Article 371. https://doi.org/10.1007/s42979-022-01276-w

Oram, F., Kapar, M. D., Saharon, A. R., Elahan, H., Segaran, P., Poloi, S., Saidal, H., Abulani, A., Lackman, I., & Ancrenaz, M. (2022). Engaging the enemy: Orangutan (*Pongo pygmaeus morio*) conservation in human modified environments in the Kinabatangan floodplain of Sabah, Malaysian Borneo. *International Journal of Primatology*, *43*(6), 1067–1094. https://doi.org/10.1007/s10764-022-00288-w

Perepolkin, D. (2019). *polite*: Be nice on the web. Version 0.1.3. https://doi.org/10.32614/CRAN.package.polite

Qomariah, I. N., Rahmi, T., Said, Z., & Wijaya, A. (2018). Conflict between human and wild sumatran elephant (*Elephas maximus sumatranus* Temminck, 1847) in Aceh Province, Indonesia. *Biodiversitas*, *20*(1), 77–84. https://doi.org/10.13057/biodiv/d200110

Risdiyanto, I., Santosa, Y., Santoso, N., & Sunkar, A. (2023). MaxEnt model application for conflict mitigation of bornean orangutan (*Pongo pygmaeus wurmbii*) in oil palm plantation. *IOP Conference Series: Earth and Environmental Science*, *1243*, Article e012022. https://doi.org/10.1088/1755-1315/1243/1/012022

Sherman, J., Ancrenaz, M., & Meijaard, E. (2020). Shifting apes: Conservation and welfare outcomes of bornean orangutan rescue and release in Kalimantan, Indonesia. *Journal for Nature Conservation*, *55*, Article 125807. https://doi.org/10.1016/j.jnc.2020.125807

Soehartono, T., Susilo, H. D., Andayani, N., Utami-Atmoko, S. S., Sihite, J., Saleh, C., & Sutrisno, A. (2007). *Strategi rencana aksi konservasi orangutan Indonesia 2007–2017*. Jakarta: Direktorat Jenderal Perlindungan Hutan dan Konservasi Alam, Departemen Kehutanan Republik Indonesia.

Suwannarong, K., Soonthornworasiri, N., Maneekan, P., Balthip, K., Yimsamran, S., Maneewatchararangsri, S., Ponlap, T., Saengkul, C., Lantican, C., Thammasutti, K., & Singhasivanon, P. (2023). Love or conflict: A qualitative study of the human-long tailed macaque interface in Nakhon Sawan Province, Thailand. *Acta Tropica*, *240*, Article 106861. https://doi.org/10.1016/j.actatropica.2023.106861

Tala, F. Z. (2016). *A study of stemming effects on information retrieval in Bahasa Indonesia.* The Netherlands: Universiteit van Amsterdam.

Utami-Atmoko, S. S, Traylor-Holzer, K., Rifqi, M. A., Siregar, P. G., Achmad, B., Priadjati, A., Husson, S. J., Wich, S. A., Hadisiswoyo, P., Saputra, F., Campbell-Smith, G., Kuncoro, P., Russon, A., Voigh, M., Santika, T., Nowak, M., Singleton, I., Sapari, I., Meididit, A., ..., & Less, C. M. (Eds.). (2017). *Orangutan population and habitat viability assessment*: Final report. Bogor: IUCN/SSC Conservation Breeding Specialist Group.

van Der Loo, M. (2013). *stringdist*: Approximate string matching, fuzzy text search, and string distance functions. Version 0.9.15. https://doi.org/10.32614/CRAN.package.stringdist

Wibowo, A. P., & Papilaya, F. S. (2020). Analisis pola kebakaran lahan di Kalimantan Timur dengan MODIS dan VIIRS. *Media Komunikasi Geografi*, *21*(1), 84–98. https://doi.org/10.23887/mkg.v21i1.23253

Wickham, H. (2009). *stringr*: Simple, consistent wrappers for common string operations. Version 1.5.1. https://doi.org/10.32614/CRAN.package.stringr

Wickham, H. (2014). *rvest*: Easily harvest (Scrape) web pages. Version 1.0.4. https://doi.org/10.32614/CRAN.package.rvest

Wickham, H. (2016). *ggplot2*: Elegant graphics for data analysis. https://doi.org/10.1007/978-0-387-98141-3

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2015). *dplyr*: A grammar of data manipulation. https://doi.org/10.32614/CRAN.package.dplyr

Yi, L., Liu, B., & Li, X. (2003). Eliminating noisy information in Web pages for data mining. In *Proceedings of the ninth ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 296–305). Association for Computing Machinery, New York, USA. https://doi.org/10.1145/956750.956785