Original scientific paper

UDK 620.92:621.311.24(497.7)
DOI 10.7251/COMEN2502109K

ANALYZING MEDIA NARRATIVES ON RENEWABLE ENERGY IN MACEDONIA USING SENTIMENT AND TOPIC MODELING

Stefani Kulebanova[1*], Marija Stojcheva[2], Aleksandra Dedinec[3], Jana Prodanova[4], Aleksandar Dedinec[5], Ljupcho Kocarev[6]

[1] Ss. Cyril and Methodius University, Faculty of Computer Science and Engineering, Skopje, Macedonia
[2] Ss. Cyril and Methodius University, Faculty of Computer Science and Engineering, Skopje, Macedonia
[3] Ss. Cyril and Methodius University, Faculty of Computer Science and Engineering, Skopje, Macedonia
[4] Macedonian Academy of Sciences and Arts, Skopje, Macedonia
[5] Macedonian Academy of Sciences and Arts, Skopje, Macedonia
[6] Macedonian Academy of Sciences and Arts, Skopje, Macedonia

[*] Corresponding author: stefani.kulebanova@students.finki.ukim.mk

Abstract: This paper explores the evolution of media sentiment and thematic discourse around Renewable Energy Sources (RES) in Macedonia from 2017 to 2024. Drawing on over 9,500 news teasers retrieved from Time.mk, we apply FinBERT and BERTweet for sentiment analysis and Latent Dirichlet Allocation (LDA) for topic modeling. Our findings reveal a dominant neutral sentiment profile with notable positive sentiment peaks aligned with key regulatory developments and market shifts. Topic modeling further captures the transformation of discourse from centralized, policy-driven rhetoric toward decentralized, technologically mature themes, centered on solar expansion, public engagement, and institutional implementation. The outcome of this research contributes to understanding how public discourse on RES evolves over time, and its implications add value to the role of media in shaping societal perception and stakeholder's response to renewable energy transition.

Keywords: renewable energy, media discourse, sentiment analysis, topic modeling, Macedonia, FinBERT, BERTweet, natural language processing.

1. INTRODUCTION

As societies accelerate energy transitions, the media has become a powerful tool through which renewable energy developments are announced, interpreted and (dis)approved [1] [2]. Media, including both traditional platforms and social media, play a crucial role in shaping public understanding, acceptance, and criticism of RES. This role is particularly pronounced in regions such as the Western Balkans, where legacies of fossil fuel dependence, fragmented governance structures, and slow institutional reforms complicate the path toward decarbonization [3] . Research from multiple countries demonstrates that applying sentiment analysis and topic modeling to media content helps uncover shifts in public opinion and the underlying factors driving support or opposition to renewable energy.

In Germany, topic modeling of media content revealed that after the Renewable Energy Act was introduced, media coverage became predominantly optimistic about renewable energy developments [4]. Similarly, a study in Norway found that although protests and resistance to wind energy increased after 2018, Twitter reflected a more nuanced national dialogue with only a moderate rise in negative sentiment [5]. In the European Union, text mining of National Energy and Climate Plans (NECPs) identified key drivers and obstacles within the renewable energy value chain, emphasizing the value of cross-coun-

try comparisons and collaborative approaches [6]. In Indonesia, sentiment and topic analysis of both news and social media highlighted generally positive views toward photovoltaics, while also revealing persistent barriers such as misconceptions, regulatory issues, and limited government support [7]. In South Korea, social media analysis suggested that while the public may view renewable and nuclear energy as complementary, they also perceive competition between them during periods of negative discourse, an insight useful for designing integrated energy policies [8].

These international findings underscore the potential of media analysis to inform energy policy and communication strategies. Conducting similar research in Macedonia provide valuable insights into public discourse around RES and support more informed, inclusive energy transitions. Macedonia offers an exciting case for studying the communicative dynamics of RES. Following the adoption of the Energy Law in 2018 and subsequent proposals regarding change in tariffs, the country has experienced a proliferation of solar and wind energy initiatives. Yet, the extent to which these initiatives transform into solid public discourse remains under-examined.

The current study addresses that gap by employing Natural Language Processing (NLP) techniques to analyze over 9,500 media sources collected via Time.mk, implementing FinBERT and BERTweet models to classify sentiment duality in RES-related headlines and teasers. Additionally, LDA is used to extract evolving thematic discussions across the 2017–2024 timeframe. These tools allowed us to identify the tone and recognize the substantive evolution of the RES discourse, emphasizing their alignment with policy milestones, technological deployment, and social engagement. This dual-method framework contributes to the growing body of literature on the importance of media in shaping public perception of renewable energy, offering empirical evidence relevant to scholars and experts in energy policy, communication studies, and regional development sectors.

The remaining parts of the paper are organized as follows: Section 2 explains the methodology covering data collection, sentiment analysis, and topic modeling. Section 3 presents the results and the subsequent discussion. The study ends with Section 4 discussing the conclusions on policy, research implications and limitations and future research lines.

2. METHODOLOGY

2.1. Data collection and preprocessing

The dataset used in this study consists of textual news teasers collected from *Time.mk*, one of the largest and most widely used news aggregators in Macedonia. Time.mk indexes headlines and short summaries from over 120 national and local media sources, including newspapers, online portals, and TV station websites. Due to its comprehensive coverage and real-time indexing, Time.mk offers an ideal source for media trend analysis on a certain topic.

To extract discourse related to RES, a keyword-based crawling procedure was employed for the period 2017–2024. The search keywords were in Macedonian language and the full list of terms is:

- **Macedonian**: *фотонапон*, *фотоволт*, *ветерна*, *ветерни*, *ветерен*, *обновливи извори на енергија*, *ОИЕ*, *повластени производители*, *повластена тарифа*, *повластени тарифи*, *премиум тарифа*, *премиум тарифи*
- **English**: *photovoltaic(s)*, *wind*, *wind power*, *renewable energy sources*, *RES*, *preferential producer(s)*, *feed-in tariff(s)*, *premium tariff(s)*

The extracted text was filtered to remove duplicates and cleaned to standardize formatting and eliminate non-informative tokens (e.g., URLs, symbols). The final corpus was segmented by year to enable longitudinal analysis of public discourse on renewable energy. In total, over 9,500 entries were processed, yielding a robust and temporally diverse corpus suitable for annual comparative analysis.

2.2. Sentiment analysis – FinBERT and BERTweet

Given the heterogeneous nature of media discourse, ranging from formal institutional updates to emotionally charged headlines, this study adopts a dual-model approach for sentiment analysis. Specifically, FineBERT and BERTweet, both based on the Bidirectional Encoder Representation from Transformers (BERT) architecture [9] are integrated to leverage their complementary strengths and ensure robust sentiment classification across diverse textual formats.

Both models follow the same underlying architecture. Let a given text input sequence, such as teaser or article excerpt be tokenized as:

$$x = \{x_1, x_2, \ldots, x_T\},$$

where T is the sequence length and x_t is the token at position t. Each token is converted into an embedding vector using a pre-trained vocabulary embedding matrix E:

$$e_t = E(x_t) \text{ for all } t \text{ in } [1, \ldots, T]$$

These embeddings are passed through a series of transformer encoder layers. For each token position t, the model computes a contextualized representation using the self-attention mechanism:

$$h_t = TransformerEncoders(e_1, \ldots, e_T)$$

The final hidden state corresponding to the special classification token [CLS] is extracted as the aggregate representation of the entire input sequence:

$$z = h_{[CLS]}$$

A softmax layer is then applied to produce the probability distribution over sentiment classes:

$$\hat{y} = softmax(Wz + b)$$

where W and b are learnable parameters. The output $\hat{y}$ is a vector of probabilities associated with the sentiment classes:

$$\hat{y} = [P(positive), P(neutral), P(negative)]$$

The predicted sentiment label y^* corresponds to the class with the highest predicted probability:

$$y^* = arg\max_i \hat{y}_i$$

FinBERT [10] is a domain-specific adaptation of BERT, fine-tuned on financial corpora such as earnings reports, company filings, and investment news, using the Financial PhraseBank dataset. This training enables it to detect sentiment in complex and often implicit expressions typical of economic and policy-oriented discourse, where sentiment is embedded in technical terminology rather than explicit emotional language. It is particularly sensitive to lexical patterns such as "subsidy approved," "investment stalled," or "production target met," which may not convey strong sentiment in general contexts but are highly indicative of sentiment in financial and regulatory domains.

In the context of RES in Macedonia, FinBERT is well-suited for interpreting sentiment in articles discussing government RES policies (e.g., feed-in tariffs, premium tariffs), investment and funding announcements, regulatory frameworks and bank financing, private-public partnership updates etc.

BERTweet [11] is a RoBERTa-based language model pre-trained on 850 million English tweets, making it effective for analyzing short, informal, and emotionally expressive texts. It utilizes a byte-pair encoding scheme adapted to the unique characteristics of social media, allowing it to interpret slang, abbreviations, hashtags, and colloquial language, which are features commonly found in media teasers and user-generated content.

In this study, BERTweet is employed to capture public sentiment expressed in teaser headlines and socially driven narratives. Its domain-specific training enables nuanced detection of tone and emotional cues, such as enthusiasm, frustration, skepticism, or outrage, which are often embedded in informal textual formats but may be overlooked by general-purpose models.

By combining FinBERT and BERTweet, the sentiment analysis pipeline captures both formal institutional tone and informal social expression, offering a comprehensive view of sentiment dynamics across sources and time periods.

2.3. Topic modeling - LDA

LDA (Latent Dirichlet Allocation) is a probabilistic generative model used to identify latent topics in a corpus based on word co-occurrence patterns [12]. This method helps uncover the thematic structure within media coverage, providing a deeper understanding of public discourse. LDA assumes that each document is a mixture of topics, with each topic represented as a distribution over words. For a corpus of M documents, LDA aims to infer the following: for each topic k, a word distribution ϕ_k over the vocabulary, and for each document d, a topic distribution θ_d, which represents the proportion of each topic present within the document. The generative process begins by sampling a topic from θ_d, followed by sampling a word from the corresponding topic's word distribution ϕ_k. This process is governed by two Dirichlet priors, α and β, which regulate the sparsity of the document-topic and topic-word distributions, respectively.

The LDA model was implemented using the Gensim library, with each year's corpus modeled independently to capture thematic shifts over time. The number of topics K for each year was determined by maximizing topic coherence scores, supplemented by manual inspection to ensure interpretability.

Ranked lists of words associated with a given word w_n are obtained by calculating the sum of the weight of each topic generated by LDA multiplied by the weight of each word w_n contained in that topic. The ranking weight of the word i is computed as follows:

$$w_i = \sum_1^N w_{ij} * w_{nj}$$

where N is the number of topics and w_{ij} denotes the weight of the word i in topic j.

To visualize and compare thematic shifts across years, several methods were employed. A heatmap of term overlap was created to illustrate shared vocabulary and track semantic drift across topics. In addition, word clouds were generated, color-coded by year, to emphasize the most significant and consistent terms in each annual corpus. These methods provided a clear view of how topics evolved over time and how specific terms persisted or changed across the years.

3. RESULTS AND DISCUSSION

3.1. Sentiment analysis

The yearly sentiment distribution identified by FinBERT (Figure 1.) reveals a predominance of neutral sentiment across the period from 2017 to 2024. However, positive sentiment shows a noticeable increase starting in 2020, peaking in 2022, and then slightly declining in 2023 and 2024. In contrast, negative sentiment remains low, indicating limited media coverage of problems or controversies associated with RES installations.

This trend can be attributed to key policy developments and market dynamics. The adoption of the Energy Law in 2018, followed by the implementation of relevant secondary legislation in the subsequent years, and the launch of tenders for feed-in premium tariffs on both public and private land, created a favorable investment climate for RES projects. These policy actions are clearly reflected in the increasing share of positive sentiment. Moreover, the energy crisis in 2022, accompanied by high electricity prices, further enhanced the economic attractiveness of renewable energy, particularly among private sector actors, thereby reinforcing the positive tone in media coverage during that period.

Figure 2. shows sentiment trends using the BERTweet model, which captures a slightly more neutral peak. Positive sentiment is also pronounced in this model, peaking during 2022 - indicating heightened enthusiasm as RES initiatives gained visibility. A comparison between FinBERT and BERTweet reveals broadly similar outcomes, with both models indicating that the majority of RES-related news content is classified as neutral—likely reflect-

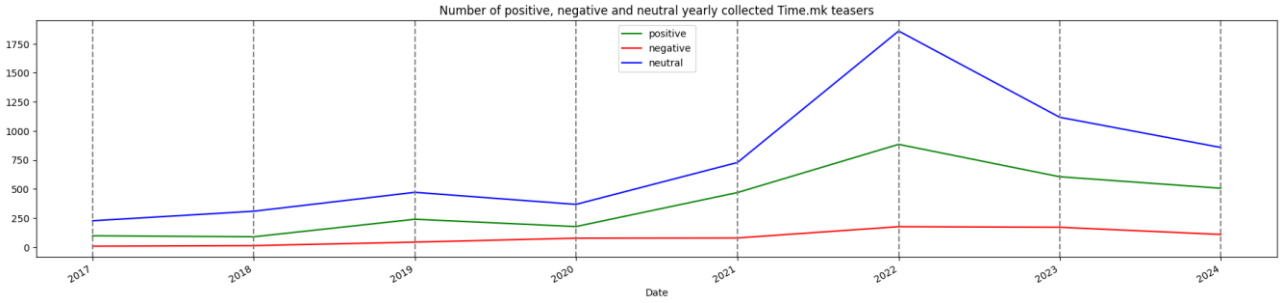

Figure 1. *Time series on the number of collected new teasers by type of sentiment, using FinBERT*

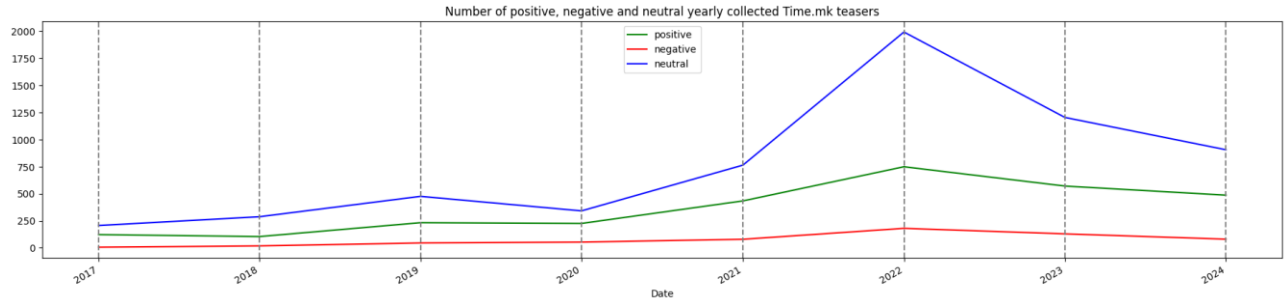

Figure 2. *Time series on the number of collected news teasers by type of sentiment, using BERTweet*

ing the predominantly factual and informative nature of media coverage on renewable energy developments.

The proportional breakdown reveals a consistent dominance of neutral sentiment in both FinBERT's (Figure 3.) and BERTweet (Figure 4.) predictions. Initially, the share of negative news was relatively low. However, as the installation of renewable RES capacities increased, the proportion of negative sentiment also rose, reaching approximately 12% in 2020. These reports mainly reflected concerns about political interference in RES development. In one case, financial authorities launched an investigation into suspected irregularities in land sales for photovoltaic projects. Another case involved allegations that individuals linked to former officials secured large-scale solar installations.

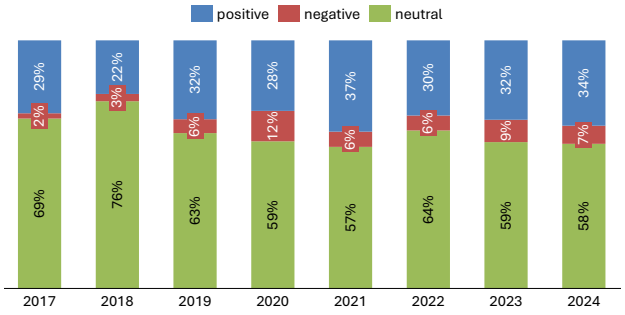

Figure 3. *Fraction of positive, negative and neutral news teasers, by year, using FinBERT*

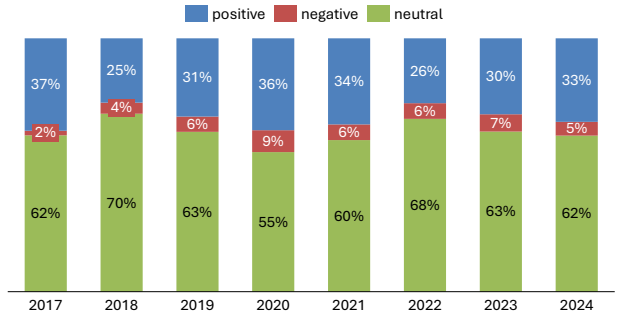

Figure 4. *Fraction of positive, negative and neutral news teasers, by year, using BERTweet*

3.2. Topic modeling

The word clouds per year provide visual cues about the most frequent and salient terms in the RES discourse, as illustrated in Figure 5, which highlights key terms like "construct", "source" and "wind" across the entire period. Figure 6. shows evolution of terminology over time, revealing that early years (2017) featured terms like "company" and "project", suggesting that the discourse during this year focused on individual renewable energy projects initiated or managed by public and private companies, signaling the utility's role as a primary driver of early investments. By 2019, the narrative appears to have focused on infrastructure implementation and institutional involvement, as indicated by the words "construct" and "ministry." This marks a growing role of government bodies (such as the Ministry of economy and the Vice President) in steering or regulating construction activities related to RES. The emphasis on "country" in 2020 suggests a more national-scale framing of renewable energy issues.

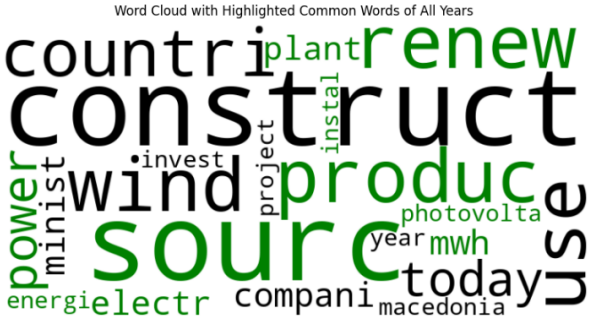

Figure 5. *Word cloud of frequent terms in the period 2017-2024 for RES discourse*

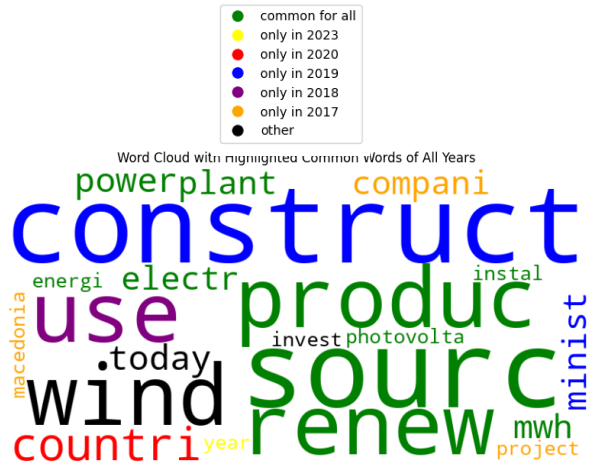

Figure 6. *Word cloud showing evolution of RES terminology by year*

Figure 7. presents the lexical continuity across years. The heatmap illustrates a clear increase in shared terminology occurs post-2018, signaling stabilization of the RES lexicon. This suggests that as policies and technologies became embedded in public discourse, the vocabulary around RES topics standardized.

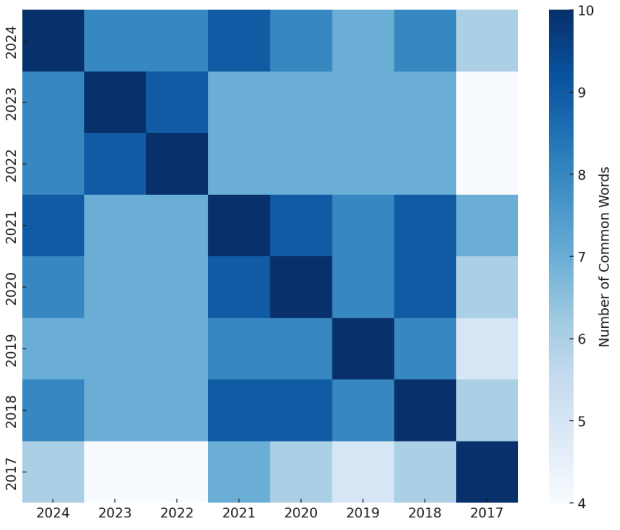

Figure 7. *Lexical continuity heatmap of RES terminology*

In addition to these analyses, we aimed to explore in greater depth the evolving discourse RES in the media over time. This analysis presents the temporal evolution of public and institutional discourse over the period 2017–2024, based on topic modeling applied to the news articles. The use of unsupervised machine learning (LDA) enables the extraction of latent thematic structures, offering insights into how the focus and framing of RES topics have shifted over time (Figure 8.):

• **2017–2018: Institutional framing -** In the initial years of the analysis (2017 and 2018), the dominant topics center on environmental concerns, government policy frameworks, and early-stage discussions on wind and solar technologies. Keywords such as "environment", "economy", "energy", "company", "wind", "solar", and "investment" suggest that discourse was predominantly shaped by macro-level narratives aimed at justifying renewable energy through climate and environmental imperatives. These years reflect a top-down orientation, with public and private companies establishing an initial individual renewable energy projects. Notably, references to "European" and "bank" also imply early engagement with external financial and policy actors, potentially linked to alignment with EU climate and energy goals.

• **2019–2020: Local integration -** From 2019 onwards, there is a noticeable shift toward localized implementation of RES deployment. The frequent appearance of terms such as "municipality", "land", "public", "system", and "building" indicates that renewable energy discourse began to engage with issues of land use planning, and public sector infrastructure. The inclusion of "partnership", "citizen", and "investment" in 2020 highlights increasing stakeholder involvement and growing public interest. This period corresponds to the decentralization of energy planning and the operationalization of national RES targets defined in the Energy Strategy adopted in 2019, possibly supported by donor-funded programs or national subsidies and public-private partnership for solar installations.

• **2021–2022: Decentralized adoption -** In 2021 and 2022, the thematic structure becomes more diversified, indicating the mainstreaming of RES discourse across multiple societal sectors. Keywords such as "household", "investment", "building", "solar", "government", and "roof" signal a transition toward residential, public and commercial and industry sector engagement, with possible expansion of distributed generation models. The term "household" reflects a marked shift to bottom-up adoption, where private actors begin to participate in the RES transition through net-billing regulation and public awareness campaigns. The appearance of "esm", "produced", "hour", and "announced" in 2022 suggests media visibility and public announcements, indicating that RES projects were becoming more tangible and widely communicated, mostly by the Energy regulatory commission of Macedonia.

• **2023–2024: Technological maturity and production emphasis -** In the most recent years (2023–2024), the topic modeling results suggest an increasing emphasis on RES production, technological components, and project execution. Keywords such as "production", "panel", "installed", "announced", "project", and "company" are indicative of a mature discourse that now centers around the implementation and commissioning of renewable energy systems. The frequent co-occurrence of "solar", "wind", "school", and "country" hints at the proliferation of photovoltaic panels in public buildings and educational institutions, led by the government and municipalities

Furthermore, to provide a broader interpretation of the results and policy implication, we outline longitudinal dynamics observed across all years of the dataset, examining recurring themes, shifts in narrative focus and temporal lags in discourse-to-deployment cycles:

Thematic continuity: Certain keywords, particularly "energy", "environment", "solar", "investment", and "government", are present throughout all years, demonstrating consistent interest and institutional commitment to the RES transition.

Increasing decentralization: There is a clear shift from centralized, government-led framing in early years to participatory, company and residential-level narratives, consistent with global trends in energy democratization.

Technology prioritization: The dominance of "solar" over "wind" in most years reflects a technological preference, likely due to solar's scalability, declining costs, reduced risk and suitability for distributed installations in Macedonia.

Temporal lag in project implementation: The evolution from abstract environmental justifications (2017–2018) to practical project rollouts (2023–2024) suggests a 5 to 7 year implementation lag from discourse to deployment—consistent with planning, permitting, and financing cycles in RES development.

These results, summarized in Figure 8. suggest an evolution aligned with the institutional and technological consolidation of the RES sector. The synchronization of narrative shifts and implementation timelines highlights the role of media in shaping the trajectory of Macedonia's energy transition.

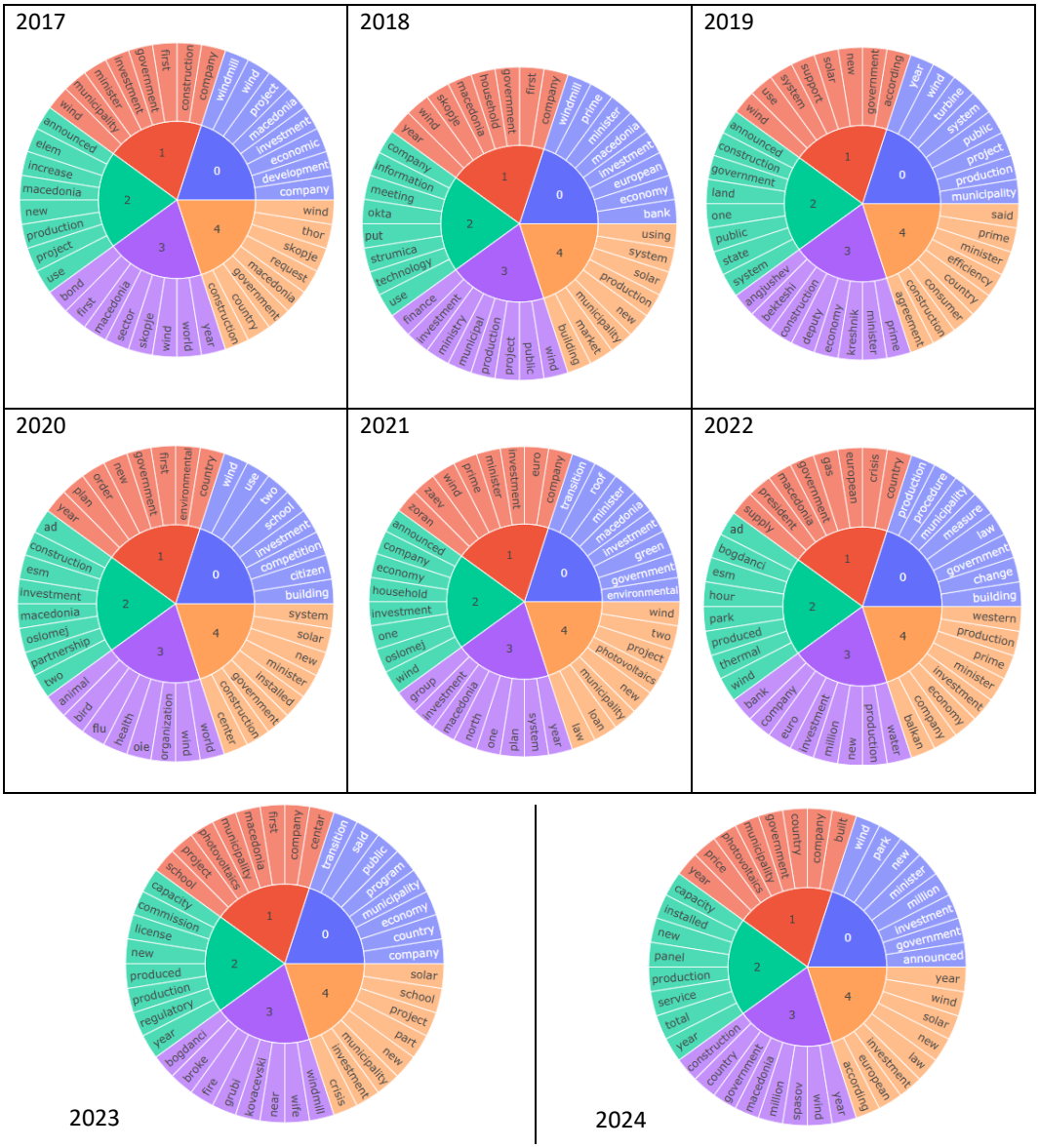

Figure 8. *LDA topic modeling of RES discourse showing evolution of thematic clusters and their prominence by year*

4. CONCLUSION

The present analysis addresses the evolving character of public discourse on topics related to renewable energy in Macedonia, shaped by regulatory triggers and media responsiveness. The sentiment analysis indicates a consistently neutral tone with brief boost of optimism, particularly during periods of policy innovation, such as after the adoption of the Energy Law in 2018, and market turbulence like the energy price increase in 2022. This trend was captured by both FinBERT and BERTweet, although showing slight differences in emotive sensitivity due to their distinct training [10] [11].

Moreover, the topics identified through LDA show a shift over time from high-level institutional language to more practical, implementation-focused communication. Early themes emphasized "government," "investment," and "policy," while later periods prioritized terms such as "solar panel," "school," and "production," which indicates noticeable organizational and decentralization efforts. This shift aligns with the broader European patterns of energy democratization and reflects the 5 to 7 year lag between discourse formation and material implementation.

This study's findings offer important implications for public and private companies in the sector. First, it is important to note that media serves as a channel for communication of the current situation in RES application, progress and future. It can be easily employed as a path for interpretation and contextualization of policies and regulatory frameworks. Second, sentiment dynamics have shown to be a critical approach to anticipate public awareness, perceptions and receptivity to future reforms. Lastly, the more stable language use after 2019 suggests a more mature policy discourse, with important indicators for how stakeholders communicate and gain support.

To end, there are several limitations that open space for future research. First, the analysis focused only on headlines and teasers, which might not reflect the depth of the articles. Future work could include full-text analysis to capture more detailed sentiment and narrative structure.

Second, the study was limited to Macedonia, restricting broader regional conclusions. This issue can be overcome with additional analysis to other Balkan countries, which would provide comparative insights into how different media systems present renewable energy transitions.

Third, while the current topic modeling approach relied solely on LDA, future research will incorporate additional methods—such as BERTopic—to improve semantic richness and capture contextual relationships more effectively.

5. REFERENCES

[1] S. Shrestha, I. Bittencourt, A. S. Varde and P. Lal, "AI-Based Modeling for Textual Data on Solar Policies in Smart Energy Applications," 2024 *15th International Conference on Information, Intelligence, Systems & Applications (IISA),* pp. 1–8, 2024.

[2] A. E. Esiri, J. M. Kwakye, D. E. Ekechukwu and O. Benjamin, "Public perception and policy development in the transition to renewable energy," *Magna Scientia Advanced Research and Reviews*, vol. 8, no. 2, pp. 228–237, 2023.

[3] J. Ignjatović, S. Filipović and M. Radovanović, "Challenges of the green transition for the recovery of the Western Balkans," *Energy, Sustainability and Society*, vol. 14, no. 2, 2024.

[4] J. Dehler-Holland, K. Schumacher and . W. Fichtner, "Topic modeling uncovers shifts in media framing of the German renewable energy act," *Patterns*, vol. 2, no. 1, 2021.

[5] O. Vågerö, A. Bråte, A. Wittemann, J. Y. Robinson, M. Zeyringer and N. Sirotko-Sibirskaya, "Machine learning of public sentiments towards wind energy in Norway," *Wind Energy*, vol. 27, no. 6, pp. 583–611, 2024.

[6] M. Kumar and J. Ng, "Using text mining and topic modelling to understand success and growth factors in global renewable energy projects," *Renewable Energy Focus*, vol. 42, pp. 211–220, 2022.

[7] Y. P. Mulyani, A. Saifurrahman, . H. M. Arini, A. Rizqiawan, B. Hartono, D. S. Utomo , A. Spanellis, M. Beltran, K. M. Nahor, D. Paramita and W. Harefa, "Analyzing public discourse on photovoltaic (PV) adoption in Indonesia: A topic-based sentiment analysis of news articles and social media," *Journal of Cleaner Production*, vol. 434, p. 140233, 2024.

[8] D. Jeong, H. Syjung, J. Kim, H. Yu and E. Park, "Public perspective on renewable and other energy resources: Evidence from social media big data and sentiment analysis," *Energy Strategy Reviews* , vol. 50, p. 101243, 2023.

[9] J. Devlin, M.-W. Chang, K. Lee and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, vol. 1, pp. 4171–4186, 2019.

[10] D. Araci, *Finbert: Financial sentiment analysis with pre-trained language models*., arXiv, 2019.

[11] D. Q. Nguyen, T. Vu and . A. T. Nguyen, *BERTweet: A pre-trained language model for English Tweets*, arXiv, 2020.

[12] D. M. Blei, A. Y. Ng and M. I. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, vol. 3, pp. 992–1022, 2003.

АНАЛИЗА МЕДИЈСКИХ НАРАТИВА О ОБНОВЉИВИМ ИЗВОРИМА ЕНЕРГИЈЕ У МАКЕДОНИЈИ ПРИМЈЕНОМ АНАЛИЗЕ СЕНТИМЕНТА И МОДЕЛОВАЊА ТЕМА

Сажетак: У овом раду испитује се еволуција медијског сентимента и тематског дискурса о обновљивим изворима енергије (ОИЕ) у Македонији у периоду од 2017. до 2024. године. Анализа се заснива на више од 9.500 новинских најава преузетих са портала Time.mk, при чему су за анализу сентимента примијењени модели FinBERT и BERTweet, док је за моделовање тема коришћена метода латентне Дирихлеове алокације (LDA). Резултати указују на доминантно неутралан сентимент, уз изражене позитивне пикове који се временски подударају са кључним регулаторним промјенама и тржишним помацима. Моделовање тема додатно осветљава трансформацију дискурса – од централизоване, политикама вођене реторике ка децентрализованим и технолошки зрелијим темама, са фокусом на ширење соларне енергије, учешће јавности и институционалну имплементацију. Добијени налази доприносе бољем разумијевању начина на који се јавни дискурс о обновљивим изворима енергије мијења током времена, те наглашавају значај улоге медија у обликовању друштвене перцепције и реакција релевантних актера у процесу енергетске транзиције.

Кључне ријечи: обновљиви извори енергије, медијски дискурс, анализа сентимента, моделовање тема, Македонија, FinBERT, BERTweet, обрада природног језика.

Paper received: 16 May 2025 **Paper accepted: 17 November 2025**