

ABSTRACT

English-language news coverage plays an important communicative role in shaping international understanding of Indonesia's energy transition, yet this domain remains under-analyzed through computational lenses. This study applies a methodologically rigorous Latent Dirichlet Allocation (LDA) pipeline to identify and interpret major themes in English-language news coverage of Indonesia's energy transition. Addressing common inconsistencies in existing LDA applications, this research emphasizes an enriched preprocessing pipeline alongside systematic hyperparameter tuning. Model quality was assessed using an evaluation framework combining quantitative metrics and qualitative validation.

The results demonstrate that preprocessing design plays a more substantial role in improving topic quality than hyperparameter variation alone. The final model extracted 16 topics, which were interpreted into six thematic clusters: financing and carbon markets, renewable energy infrastructure, biofuels and resource strategy, international energy trade and geopolitics, emerging low-carbon technologies, and urban and nature-based climate initiatives. The findings suggest that English-language coverage predominantly frames Indonesia's energy transition through financial, diplomatic, and geopolitical narratives. These results offer descriptive insight into English-language news framing surrounding Indonesia's transition and demonstrate the value of an enhanced topic modeling pipeline as an analytical tool for understanding dominant narratives within large-scale domain-specific corpora.

Keywords: LDA, topic modeling, Indonesia, energy transition, English-language news

ABSTRAK

Pemberitaan berbahasa Inggris memiliki peran komunikatif yang penting dalam membentuk pemahaman internasional mengenai transisi energi Indonesia, namun domain ini masih relatif jarang dianalisis melalui pendekatan komputasional. Penelitian ini menerapkan *Latent Dirichlet Allocation* (LDA) dengan alur kerja yang dirancang secara metodologis dan sistematis untuk mengidentifikasi serta menginterpretasikan tema-tema utama dalam pemberitaan berbahasa Inggris mengenai transisi energi Indonesia.

Sebagai respons terhadap berbagai inkonsistensi umum dalam penerapan LDA pada penelitian sebelumnya, studi ini menekankan penggunaan alur praproses yang diperkaya serta pemilihan *hyperparameter* terbaik. Kualitas model dievaluasi menggunakan kerangka evaluasi yang menggabungkan metrik kuantitatif dan validasi kualitatif.

Hasil penelitian menunjukkan bahwa desain praproses memiliki pengaruh yang lebih signifikan terhadap peningkatan kualitas topik dibandingkan variasi hiperparameter semata. Model akhir berhasil mengekstraksi 16 topik, yang kemudian diinterpretasikan ke dalam enam kluster tematik: pembiayaan dan pasar karbon, infrastruktur energi terbarukan, biofuel dan strategi sumber daya, perdagangan energi internasional dan geopolitik, teknologi rendah karbon yang berkembang, serta inisiatif iklim berbasis perkotaan dan alam.

Temuan ini mengindikasikan bahwa pemberitaan berbahasa Inggris cenderung membingkai transisi energi Indonesia melalui narasi finansial, diplomatik, dan geopolitik. Hasil penelitian ini memberikan gambaran deskriptif mengenai pola pembedaan media berbahasa Inggris terhadap transisi energi Indonesia serta menunjukkan nilai tambah dari pipeline pemodelan topik yang ditingkatkan sebagai alat analitis untuk memahami narasi dominan dalam korpus berskala besar dan spesifik domain.

Kata kunci: LDA, pemodelan topik, Indonesia, transisi energi, berita berbahasa Inggris