

Abstrak

Perkembangan teknologi informasi telah menyebabkan meningkatnya jumlah artikel berita politik yang dipublikasikan secara *online*, khususnya melalui portal Detik.com. Kondisi ini memunculkan tantangan baru, yaitu bagaimana mengelompokkan dan memahami isi berita dalam jumlah besar secara efisien tanpa harus membaca satu per satu. Permasalahan utama terletak pada proses identifikasi topik yang dilakukan secara manual, yang memakan waktu lama dan tidak efektif dalam menangkap tren isu secara menyeluruh. Untuk menjawab permasalahan tersebut, penelitian ini menggunakan metode *Latent Dirichlet Allocation (LDA)* untuk melakukan pemodelan topik pada artikel berita politik dalam rentang waktu Januari 2021 hingga Mei 2025. Proses pemodelan diawali dengan tahapan preprocessing teks seperti *cleaning*, *stopword removal*, *normalisasi*, *tokenisasi*, hingga pembentukan *bag-of-words*. Penelitian ini bertujuan untuk menganalisis pengaruh parameter pemodelan seperti jumlah topik, *iterasi (passes)*, dan kualitas preprocessing terhadap hasil topik yang dihasilkan oleh model *LDA*. Hasil eksperimen menunjukkan bahwa konfigurasi terbaik diperoleh pada jumlah topik sebanyak 13 dan iterasi sebanyak 5, dengan nilai *coherence score* tertinggi sebesar 0.5250 dan *perplexity* sebesar 545,36. Nilai ini tergolong dalam kategori koherensi sedang hingga baik, yang berarti topik yang terbentuk cukup relevan dan mudah diinterpretasikan secara tematik. Pengujian dilakukan dengan variasi jumlah topik antara 2 hingga 15 dan jumlah iterasi sebanyak 4 kali, yaitu 5, 10, 15, dan 20. Peningkatan nilai *coherence score* secara bertahap diamati, dimulai dari 0.3374 dengan 2 topik dan 5 iterasi hingga mencapai nilai tertinggi 0.5250 dengan 13 topik dan 5 iterasi. Jumlah topik yang terlalu sedikit cenderung menghasilkan topik yang terlalu umum, sedangkan jumlah yang terlalu banyak menyebabkan tumpang tindih antar topik. Di sisi lain, tahapan preprocessing terbukti sangat penting dalam menyaring kata-kata tidak relevan dan meningkatkan kualitas distribusi kata dalam tiap topik. Dengan demikian, pemilihan parameter yang tepat berkontribusi besar dalam meningkatkan keberhasilan pemodelan topik terhadap data berita politik.

Kata Kunci: *Latent Dirichlet Allocation, Pemodelan Topik, Berita Politik, Detik.com, Coherence score, dan Analisis Trend.*

Abstract

The development of information technology has led to an increase in the number of political news articles published online, especially through the portal Detik.com. This situation presents a new challenge, namely how to group and understand the content of a large amount of news efficiently without having to read each one individually. The main problem lies in the manual topic identification process, which is time-consuming and ineffective in capturing trends in issues comprehensively. To address this problem, this research uses the Latent Dirichlet Allocation (LDA) method to model topics in political news articles from January 2021 to May 2025. The modeling process begins with the stages of text preprocessing such as cleaning, stopword removal, normalization, tokenization, to the formation of a bag-of-words. This study aims to analyze the effects of modeling parameters such as the number of topics, iterations (passes), and the quality of preprocessing on the topic results generated by the LDA model. The experimental results show that the best configuration is obtained with a total of 13 topics and 5 iterations, with the highest coherence score of 0.5250 and a perplexity of 545.36. This value falls into the category of moderate to good coherence, meaning that the formed topics are quite relevant and easy to interpret thematically. The testing was conducted with a variation of the number of topics ranging from 2 to 15 and the number of iterations was 4 times, namely 5, 10, 15, and 20. A gradual increase in the coherence score was observed, starting from 0.3374 with 2 topics and 5 iterations, reaching the highest value of 0.5250 with 13 topics and 5 iterations. A very small number of topics tends to produce overly general topics, while too many topics cause overlaps between them. On the other hand, the preprocessing stage has proven to be very important in filtering out irrelevant words and improving the quality of word distribution within each topic. Thus, the selection of appropriate parameters significantly contributes to the success of topic modeling on political news data.

Keywords: *Latent Dirichlet Allocation, Topic Modeling, Political News, Detik.com, Coherence score, and Trend Analysis.*