ABSTRAK

Penelitian ini menganalisis topik pada transkrip video pidato politik keterangan pers Presiden Joko Widodo periode 2014–2024 menggunakan metode Latent Dirichlet Allocation (LDA). Tujuannya adalah mengidentifikasi perubahan distribusi frekuensi topik utama (Pendidikan, Ekonomi, Kesehatan, Infrastruktur, Teknologi) dan mengevaluasi akurasi model LDA. Data diambil dari 185 video YouTube Sekretariat Kabinet, diproses dengan preprocessing (pembersihan data, case folding, penghapusan stopwords, stemming, tokenisasi), dan dimodelkan dengan LDA. Hasil penelitian menunjukkan dominasi topik Ekonomi (35%) dan Infrastruktur (20%) pada periode 2014–2023, sementara tahun 2024 terjadi pergeseran ke topik Kesehatan (28%) dan Infrastruktur (25%). Evaluasi model menghasilkan coherence score 0,85 dan perplexity 11,94, mengindikasikan koherensi topik tinggi dan kemampuan prediksi baik. Penelitian membuktikan efektivitas LDA dalam analisis dinamika kebijakan politik serta integrasi teknologi speech-to-text untuk ekstraksi teks otomatis.

Kata kunci: Pidato Politik, Transkrip Video, Latent Dirichlet Allocation, Topic Modeling, Speech-to-Text

ABSTRACT

This research analyzes topics in press conference speech transcripts of President Joko Widodo (2014–2024) using the Latent Dirichlet Allocation (LDA) method. The objectives are to identify changes in the frequency distribution of main topics (Education, Economy, Health, Infrastructure, Technology) and evaluate LDA model accuracy. Data from 185 YouTube videos (Secretariat of the Cabinet channel) were preprocessed (cleaning, case folding, stopword removal, stemming, tokenization) and modeled with LDA. Results show the dominance of Economy (35%) and Infrastructure (20%) topics during 2014–2023, while 2024 saw a shift toward Health (28%) and Infrastructure (25%). Model evaluation yielded a coherence score of 0.85 and perplexity of 11.94, indicating high topic coherence and robust predictive ability. The study proves LDA's effectiveness in revealing policy dynamics and the utility of speech-to-text technology for automated text extraction.

Keywords: *Political Speech, Video Transcript, Latent Dirichlet Allocation, Topic Modeling, Speech-to-Text*