

ABSTRAK

Penelitian ini bertujuan untuk memetakan tren dan pola topik kekerasan remaja dalam pemberitaan *online* di Indonesia. Dengan menganalisis data artikel berita dari berbagai portal terkemuka selama periode Agustus 2024 hingga Mei 2025, penelitian ini menggunakan pendekatan kuantitatif komputasional. Proses pengumpulan data dilakukan melalui *web scraping* menggunakan *framework* Scrapy, menghasilkan total 5272 artikel yang relevan. Data teks kemudian melalui tahapan pra-pemrosesan yang meliputi *case folding*, penghapusan URL dan tag HTML, *tokenisasi*, *stopword removal*, dan *stemming* untuk menormalisasi teks. Pemodelan Topik Latent Dirichlet Allocation (LDA) berhasil mengidentifikasi 10 topik utama terkait kekerasan remaja, meliputi isu kriminalitas jalanan, kekerasan dalam rumah tangga, kekerasan seksual, serta peran hukum dan pendidikan. Analisis deret waktu menunjukkan fluktuasi signifikan dalam volume pemberitaan, dengan puncak-puncak tertentu yang mengindikasikan peningkatan perhatian media pada topik spesifik. Selanjutnya, *Association Rule Mining (ARM)*, baik berbasis kata maupun topik, mengungkap pola-pola asosiasi yang kuat. ARM berbasis topik menyoroti hubungan antara tema-tema besar seperti asosiasi antara "Kekerasan Seksual & Pencabulan" dengan "Perlindungan & Tumbuh Kembang Anak/Remaja", sementara ARM berbasis kata memberikan detail granular tentang kata kunci yang sering muncul bersama, seperti "polisi", "tangkap", dan "pelaku". Sintesis temuan mengindikasikan adanya narasi dominan yang berfokus pada insiden kekerasan dan respons institusional. Penelitian ini memberikan wawasan komprehensif mengenai dinamika pemberitaan kekerasan remaja di media *online* Indonesia, yang dapat menjadi dasar bagi pembuat kebijakan dan praktisi terkait.

Kata Kunci: Kekerasan Remaja, Pemberitaan Online, Web Scraping, Pemodelan Topik, Latent Dirichlet Allocation (LDA), Association Rule Mining (ARM), Analisis Deret Waktu (TSA).

ABSTRACT

This study aims to map the trends and patterns of youth violence topics in online news coverage in Indonesia. By analyzing news article data from various leading online portals during the period of August 2024 to May 2025, the research adopts a computational quantitative approach. Data collection was conducted through web scraping using the Scrapy framework, resulting in a total of 5,272 relevant articles. The text data underwent several preprocessing stages, including case folding, URL and HTML tag removal, tokenization, stopword removal, and stemming for text normalization. Latent Dirichlet Allocation (LDA) topic modeling successfully identified 10 main topics related to youth violence, covering issues such as street crime, domestic violence, sexual violence, and the roles of law and education. Time series analysis revealed significant fluctuations in news volume, with certain peaks indicating heightened media attention on specific issues. Furthermore, Association Rule Mining (ARM), both word-based and topic-based, uncovered strong association patterns. Topic-based ARM highlighted connections between major themes, such as the association between "Sexual Violence & Molestation" and "Child/Youth Protection & Development", while word-based ARM provided granular insights into frequently co-occurring keywords such as "police", "arrest", and "perpetrator". The synthesis of findings indicates a dominant narrative focusing on violent incidents and institutional responses. This study offers a comprehensive overview of the dynamics of youth violence reporting in Indonesian online media, serving as a valuable reference for policymakers and relevant practitioners.

Keywords: *Youth Violence, Online News Coverage, Web Scraping, Topic Modeling, Latent Dirichlet Allocation (LDA), Association Rule Mining (ARM), Time Series Analysis (TSA).*