

Komparasi Latent Dirichlet Allocation dan *BERTopic* *Modeling* dalam Identifikasi Topik pada ulasan KPU di X

Muhammad Rifqi Apriansyah

Di bawah bimbingan Dine Tiara Kusuma., ST., M.Kom

ABSTRAK

Media sosial X menjadi sarana utama masyarakat dalam menyampaikan opini terhadap kinerja Komisi Pemilihan Umum (KPU), khususnya pada konteks Pemilu 2024. Banyaknya data teks yang bersifat tidak terstruktur menuntut penggunaan metode pemodelan topik untuk mengidentifikasi isu-isu utama yang berkembang. Penelitian ini bertujuan untuk mengidentifikasi topik percakapan publik mengenai KPU di X serta membandingkan kinerja metode Latent Dirichlet Allocation (*LDA*) dan *BERTopic* dalam pemodelan topik. Metode penelitian meliputi pengumpulan data komentar terkait KPU, tahap prapemrosesan teks, pemodelan topik menggunakan *LDA* dan *BERTopic*, serta evaluasi kualitas topik menggunakan nilai koherensi dan interpretabilitas. Hasil penelitian menunjukkan bahwa *BERTopic* menghasilkan topik yang lebih koheren dan kontekstual dibandingkan *LDA* karena memanfaatkan representasi semantik berbasis *embedding* transformer. Sementara itu, *LDA* mampu memodelkan distribusi kata secara probabilistik namun cenderung menghasilkan topik yang lebih umum pada teks pendek. Dengan demikian, *BERTopic* lebih efektif dalam mengidentifikasi topik berbasis makna pada ulasan KPU di X, sedangkan *LDA* lebih sesuai untuk analisis distribusi kata. Penelitian ini diharapkan dapat memberikan gambaran isu publik dominan serta menjadi referensi pemilihan metode pemodelan topik pada data media sosial.

Kata kunci: *Topic Modeling*, Latent Dirichlet Allocation, *BERTopic*, Media Sosial X, KPU.

***Comparison of Latent Dirichlet Allocation and BERTopic
Modeling in Topic Identification in the
KPU Review in X***

Muhammad Rifqi Apriansyah

Under the Guidance Dine Tiara Kusuma., ST., M.Kom

ABSTRACT

Social media platform X has become a major medium for the public to express opinions regarding the performance of the General Election Commission (KPU), particularly in the context of the 2024 General Election. The large volume of unstructured textual data requires the application of Topic Modeling methods to identify the main issues discussed by the public. This study aims to identify dominant topics in public discussions about KPU on X and to compare the performance of Latent Dirichlet Allocation (LDA) and BERTopic in Topic Modeling. The research methodology includes data collection of KPU-related comments, text preprocessing, Topic Modeling using LDA and BERTopic, and evaluation of topic quality based on coherence value and interpretability. The results show that BERTopic produces more coherent and contextually meaningful topics than LDA by utilizing transformer-based semantic embeddings. Meanwhile, LDA is effective in modeling word distributions probabilistically but tends to generate more general topics when applied to short texts. Therefore, BERTopic is more suitable for semantic-based topic identification in KPU-related reviews on X, while LDA is more appropriate for word distribution analysis. This study is expected to provide insights into dominant public issues and serve as a reference for selecting Topic Modeling methods for social media text analysis.

Keywords: *Topic Modeling, Latent Dirichlet Allocation, BERTopic, Social Media X, KPU*