

ABSTRACT

IMPLEMENTATION NER-LDA BASED TRANSFORMER MODEL FOR DETECTION AND CLASSIFICATION OF PUBLIC POLICY TWEETS IN INDONESIA

By

Dina Merlia Efendi

The integration of technological innovation has significantly transformed communication structures and shifted daily life toward digitalization. One of the most popular social media platforms in Indonesia is X (Twitter). Tweets can serve as valuable information for the government and stakeholders to optimize public policy. Therefore, an effective method is needed to detect and classify public concerns expressed in trending policy-related tweets on X. This study applies Named Entity Recognition – Latent Dirichlet Allocation (NER-LDA) to optimize text classification using Transformer-based models, including Indonesian Bidirectional Encoder Representations from Transformers (IndoBERT), Distilled Bidirectional Encoder Representations from Transformers Multilingual (DistilBERT Multilingual), and Cross-lingual Language Model-Robustly Optimized Bidirectional Encoder Representations From Transformers Pretraining Approach (XLM-RoBERTa). The study also implements data augmentation using Large Language Model Meta AI (LLaMA) and hyperparameter tuning using Grid Search. The best parameter combination for the three Transformer models during trending data was a batch size of 16 and a learning rate of 1×10^{-5} . For post trending data, the optimal batch size was 32 for IndoBERT and 16 for DistilBERT Multilingual and XLM-RoBERTa, with the same optimal learning rate of 1×10^{-5} across all models. The comparison of the three Transformer models shows that the best evaluation result for during trending data was achieved by the DistilBERT Multilingual model, while for post trending data was achieved by IndoBERT, with both models reaching an accuracy of 99%. These findings demonstrate that Transformer-based models perform optimally in classifying public policy tweets, both during and post trending periods.

Keywords: DistilBERT Multilingual, IndoBERT, NER-LDA, Transformer, XLM-RoBERTa.

ABSTRAK

PENERAPAN NER-LDA BERBASIS MODEL *TRANSFORMER* UNTUK DETEKSI DAN KLASIFIKASI *TWEET* KEBIJAKAN PUBLIK DI INDONESIA

Oleh

Dina Merlia Efendi

Integrasi inovasi teknologi signifikan mengubah tatanan komunikasi mentransformasikan kehidupan ke arah digital. Platform media sosial populer di Indonesia salah satunya X (Twitter). Ruang lingkup *tweet* dapat menjadi informasi sebagai acuan pemerintah dan pemangku kepentingan mengoptimalkan kebijakan di Indonesia. Oleh karena itu, diperlukan cara yang tepat untuk mendeteksi dan mengklasifikasi *tweet* keresahan masyarakat terhadap fenomena kebijakan publik yang *trending* pada platform X. Penelitian ini menerapkan *Named Entity Recognition – Latent Dirichlet Allocation* (NER-LDA) dalam mengoptimalkan klasifikasi teks menggunakan model *Transformer* yang dibandingkan mulai dari *Indonesian Bidirectional Encoder Representations from Transformers* (IndoBERT), *Distilled Bidirectional Encoder Representations from Transformers Multilingual* (DistilBERT Multilingual), hingga *Cross-lingual Language Model-Robustly Optimized Bidirectional Encoder Representations From Transformers Pretraining Approach* (XLM-RoBERTa). Penelitian ini menerapkan augmentasi data *Large Language Model Meta AI* (LLaMA) dan *hyperparameter tuning Grid Search*. Hasil kombinasi parameter terbaik pada ketiga model *Transformer* dari data saat *trending* adalah *batch size* 16 dan *learning rate* sebesar 1×10^{-5} lalu pada data setelah *trending* adalah *batch size* 32 untuk model IndoBERT serta *batch size* 16 untuk model DistilBERT Multilingual dan XLM-RoBERTa dengan *learning rate* terbaik untuk ketiga model sebesar 1×10^{-5} . Perbandingan ketiga model *Transformer* memperoleh hasil evaluasi terbaik untuk data saat *trending* adalah model DistilBERT Multilingual dan data setelah *trending* adalah model IndoBERT dengan *accuracy* kedua model mencapai 99%. Hasil menunjukkan bahwa model *Transformer* memiliki kinerja optimal dalam klasifikasi *tweet* kebijakan publik, baik pada data saat maupun setelah *trending*.

Kata kunci: DistilBERT Multilingual, IndoBERT, NER -LDA, Transformer, XLM-RoBERTa.