

INTISARI

POLITICAL MICROTARGETING DAN KLASIFIKASI PEMILIH MASYARAKAT DIGITAL INDONESIA DENGAN MENGGUNAKAN NATURAL LANGUAGE PROCESSING

Oleh

Risad Tristanto

22/502198/PPA/06423

Penelitian ini menggunakan teknik *Natural Language Processing* (NLP) untuk mengembangkan *political microtargeting* yang efektif di masyarakat digital Indonesia. Fokus utama penelitian adalah mengidentifikasi dan mengevaluasi algoritma serta teknik NLP dalam menganalisis konten politik di media sosial Indonesia, dengan mempertimbangkan karakteristik linguistik dan budaya, seperti variasi bahasa, penggunaan slang, dan data tidak terstruktur. Analisis sentimen berbasis *lexicon* menggunakan *InSet lexicon* pada pengujian awal menghasilkan akurasi sebesar 0,35, dengan *precision*, *recall*, dan *f1-score* untuk kategori negatif masing-masing bernilai 0,00 yang menunjukkan kesulitan dalam mengenali sentimen negatif. Setelah dilakukan modifikasi dengan penambahan *lexicon* lainnya dan *Swear Words List*, akurasi meningkat menjadi 0,73, dengan *precision* 0,75, *recall* 0,89, dan *f1-score* 0,82 untuk sentimen negatif. Teknik *Latent Dirichlet Allocation* (LDA) digunakan untuk mengidentifikasi sembilan topik utama dengan nilai koherensi tertinggi 0,391950 pada iterasi ke-9, yang kemudian diklasifikasikan ke dalam kategori seperti *Interest*, *Values*, *Attitudes*, dan *Lifestyles*. Dalam prediksi *voting intention*, model *Support Vector Machine* (SVM) menunjukkan performa terbaik dengan akurasi 0,76127 dan AUC 0,75934 dibandingkan dengan *regularized logistic regression* dan *artificial neural network* (ANN). Hasil klasifikasi menunjukkan bahwa pengguna Twitter dapat dikelompokkan ke dalam enam kategori utama berdasarkan dukungan politik dan niat memilih, yang menyediakan dasar untuk strategi komunikasi politik yang lebih efektif. Penelitian ini menunjukkan bahwa teknik NLP dapat mengatasi tantangan teknis dalam *political microtargeting* di Indonesia dan memberikan wawasan penting tentang karakteristik pemilih di era digital.

Kata Kunci: *Natural Language Processing, political microtargeting, analisis sentimen, LDA, support vector machine*

ABSTRACT

POLITICAL MICROTARGETING AND VOTER SEGMENTATION OF INDONESIAN DIGITAL SOCIETY USING NATURAL LANGUAGE PROCESSING

by

Risad Tristanto

22/502198/PPA/06423

This research examines the application of Natural Language Processing (NLP) techniques to develop effective political microtargeting in Indonesia's digital society. The main focus of the research is to identify and evaluate NLP algorithms and techniques in analyzing political content on Indonesian social media, taking into account linguistic and cultural characteristics, such as language variation, slang usage, and unstructured data. Lexicon-based sentiment analysis using the InSet lexicon in the initial test resulted in an accuracy of 0.35, with precision, recall, and f1-score for the negative category each valued at 0.00, indicating difficulty in recognizing negative sentiment. After modification with the addition of other lexicons and Swear Words List, the accuracy increased to 0.73, with precision 0.75, recall 0.89, and f1-score 0.82 for negative sentiment. Latent Dirichlet Allocation (LDA) technique was used to identify nine main topics with the highest coherence value of 0.391950 at the 9th iteration, which were then classified into categories such as Interest, Values, Attitudes, and Lifestyles. In predicting voting intention, the Support Vector Machine (SVM) model performed best with an accuracy of 0.76127 and AUC of 0.75934 compared to regularized logistic regression and artificial neural network (ANN). The classification results show that Twitter users can be grouped into six main categories based on political support and voting intention, which provides a basis for more effective political communication strategies. This research shows that NLP techniques can overcome the technical challenges of political microtargeting in Indonesia and provide important insights into the characteristics of voters in the digital age.

Keyword: Natural Language Processing, political microtargeting, sentiment analysis LDA, support vector machine