

ANALISIS SENTIMEN PADA PEMILIHAN PRESIDEN 2024 DI INDONESIA MENGGUNAKAN WORD2VEC DAN GATED RECURRENT UNIT (GRU) DENGAN OPTIMASI ALGORITMA GENETIKA

ABSTRACT

This study builds a sentiment analysis model for tweets related to the presidential election in Indonesia, using 39,791 data points. The model utilizes the GRU method, TF-IDF feature extraction, Word2Vec feature expansion, and Genetic Algorithm optimization. The results show an accuracy of 86.46%, which is a 4.49% improvement compared to the baseline. This performance is achieved by combining TF-IDF with a maximum of 5,000 features, applying Word2Vec with top 1 similarity, and implementing the Genetic Algorithm for feature optimization. The research demonstrates a connection between using Word2Vec feature expansion and Genetic Algorithm optimization in improving the model's accuracy.



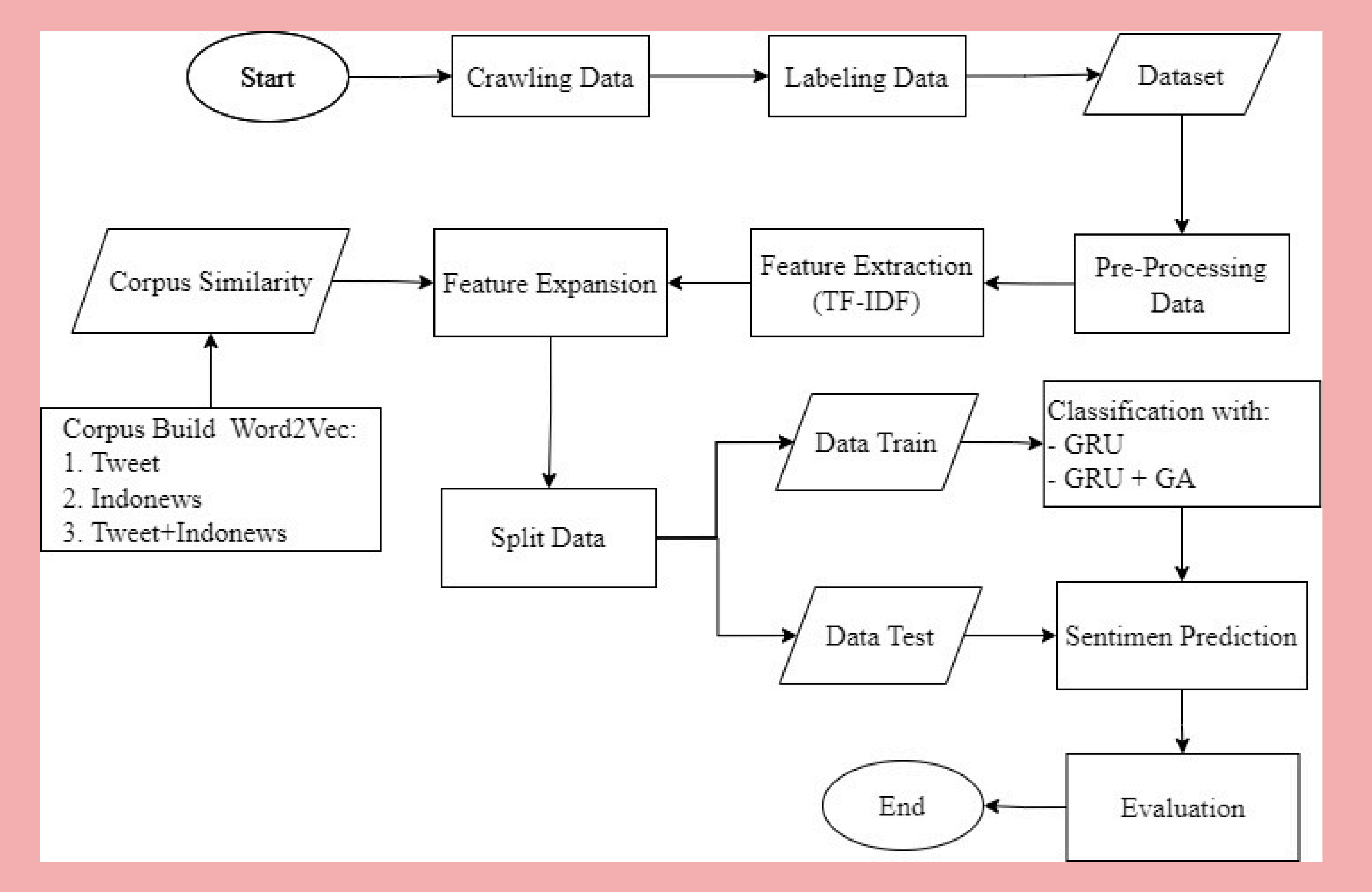
LATAR BELAKANG

Teknologi informasi telah mengubah peran media sosial menjadi platform untuk menyampaikan pendapat dan aspirasi. Twitter, dengan 19,5 juta pengguna di Indonesia, menjadi platform penting untuk menganalisis sentimen publik terhadap pemilihan presiden 2024. Penelitian ini menggunakan model Gated Recurrent Unit (GRU) yang dioptimalkan dengan Genetic Algorithm (GA) dan Word2Vec untuk menganalisis sentimen publik terhadap topik pemilihan presiden 2024 di Indonesia. Model ini akan membantu memahami persepsi dan respons masyarakat terhadap kandidat dan isu-isu politik, serta memberikan wawasan berharga bagi para pemimpin politik, kandidat, dan pemilih dalam menyusun strategi kampanye dan kebijakan.

TOPIK DAN BATASAN

- Data yang digunakan diambil dari media sosial Twitter
- Tweet yang digunakan hanya tweet yang bahasa Indonesia dengan topik pemilihan presiden di Indonesia pada tahun 2024
- Data akan dikelompokkan menjadi dua kelas yaitu positif dan negatif.

PERANCANGAN SISTEM



EVALUASI

Penelitian ini mengevaluasi model GRU dalam menganalisis sentimen terkait pemilihan presiden di Indonesia tahun 2024 menggunakan data Twitter. Pemilihan split data yang tepat dalam skenario pertama dianggap penting sebagai baseline. Skenario kedua menunjukkan bahwa penggunaan TF-IDF dengan 5.000 fitur maksimal meningkatkan akurasi dan F1-score. Skenario ketiga menyoroti bahwa penggunaan korpus IndoNews pada similarity top 1 menghasilkan akurasi tertinggi. Skenario terakhir menunjukkan bahwa Algoritma Genetika dapat meningkatkan akurasi model dengan optimalisasi fitur.

REFERENSI

Kanojia, D., & Joshi, A. (2023). Applications and Challenges of Sentiment Analysis in Real-life Scenarios. arXiv preprint arXiv:2301.09912

KESIMPULAN

Penelitian berhasil menggunakan data Twitter untuk menganalisis sentimen terkait pemilihan presiden di Indonesia. Dengan 37.391 data yang diberi label positif/negatif, model GRU diuji dalam empat skenario. Hasil terbaik adalah 86.46% akurasi, meningkat 4.49% dari baseline, dengan kombinasi TF-IDF (5.000 fitur maksimal), Word2Vec (IndoNews, top 1 similarity), dan Algoritma Genetika untuk optimasi fitur. Penggabungan ke empat skenario tersebut terbukti dapat meningkatkan nilai akurasi.