

Abstrak

Ekstraksi peristiwa dalam bahasa dengan sumber daya terbatas, seperti bahasa Indonesia, menghadapi berbagai tantangan, terutama karena keterbatasan data yang beranotasi, kompleksitas bahasa, dan ketidakseimbangan jumlah data antar kelas. Penelitian ini bertujuan untuk mengembangkan sistem ekstraksi peristiwa yang efektif dan dapat diinterpretasikan, yang dirancang khusus untuk artikel berita Indonesia terkait kasus kekerasan seksual. Isu kekerasan seksual menjadi perhatian utama karena tingginya angka kejadian serta dampak sosial yang signifikan di Indonesia. Kasus-kasus ini sering kali disampaikan melalui narasi anekdotal atau media digital, sehingga sulit untuk dianalisis secara sistematis. Untuk mengatasi tantangan ini, penelitian ini mengusulkan model Conditional Random Fields (CRF) yang dirancang khusus untuk tugas ekstraksi peristiwa. Model ini memanfaatkan berbagai fitur, termasuk fitur leksikal, kontekstual, dan semantik, guna menangkap pola serta hubungan yang terdapat dalam data. Untuk mengatasi ketidakseimbangan kelas, teknik augmentasi data, seperti penggantian sinonim, diterapkan. Model CRF dievaluasi dengan membandingkannya terhadap pendekatan lain, termasuk sistem berbasis aturan (Rule-Based) dan CNN-BiLSTM, serta menunjukkan kinerja yang kompetitif dengan skor F1 sebesar 0.730, yang menggarisbawahi pentingnya pemilihan fitur dalam lingkungan dengan sumber daya terbatas. Meskipun augmentasi data membantu meningkatkan variasi dalam data, metode ini juga memperkenalkan *noise* yang dapat membatasi peningkatan kinerja secara keseluruhan. Hasil penelitian ini menunjukkan bahwa CRF menawarkan keseimbangan antara efisiensi dan efektivitas dalam menghadapi tantangan pemrosesan bahasa alami (NLP) dengan sumber daya terbatas. Selain itu, penelitian ini memberikan dasar bagi pengembangan metode yang lebih lanjut, khususnya pendekatan *hybrid* yang menggabungkan model statistik dengan teknik *deep learning* untuk meningkatkan kemampuan ekstraksi peristiwa dalam bahasa dan bidang yang kurang terwakili.

Kata kunci: ekstraksi peristiwa, conditional random fields, augmentasi data, kekerasan seksual
