

ABSTRAKSI

Ekstraksi Informasi Kejahatan adalah tugas untuk mengekstrak informasi penting dari teks tidak terstruktur dalam domain kejahatan. Beberapa penelitian sebelumnya tentang ekstraksi informasi kejahatan menerapkan aturan untuk melakukan ekstraksi informasi kejahatan. Mendefinisikan aturan secara manual, bagaimanapun, adalah tugas yang sangat menakutkan dan kinerjanya biasanya tidak cukup baik, di mana beberapa contoh yang salah diklasifikasikan dapat ditemukan.

Sebagian besar penelitian ekstraksi informasi kejahatan telah dilakukan dalam dataset bahasa Inggris. Ada beberapa penelitian sebelumnya tentang ekstraksi informasi kejahatan yang dilakukan dalam teks bahasa Indonesia, salah satunya menggunakan ekstraksi informasi berbasis aturan dan menerapkan penandaan *Part-of-Speech (POS)* dan *Dependency Parsing* sebagai fitur. Meskipun demikian, kinerja penelitian sebelumnya harus ditingkatkan, terutama pada ekstraksi lokasi dan tanggal/waktu. Kesalahan klasifikasi disebabkan oleh sistem yang tidak dapat menentukan nama entitas dengan metode yang diusulkan pada penelitian sebelumnya.

Penelitian ini mengusulkan sebuah sistem yang mampu mengekstrak informasi kriminal pada berita online Indonesia menggunakan metode kombinasi pengenalan entitas bernama dan Support Vector Machine (SVM) untuk mengekstrak lokasi dan waktu kejahatan. Setelah lokasi dan waktu digali, dilakukan klasifikasi jenis kejahatan menggunakan metode SVM. Keluaran sistem yang diusulkan adalah: jenis kejahatan, lokasi kejahatan, dan tanggal/waktu kejadian. Evaluasi dilakukan dengan membandingkan hasil ekstraksi sistem yang diusulkan terhadap ekstraksi label emas dan hasil studi sebelumnya. Hasilnya menunjukkan bahwa sistem yang diusulkan mengungguli sistem sebelumnya secara signifikan. Meskipun kinerjanya meningkat, hasil menunjukkan metode yang diusulkan masih perlu ditingkatkan di beberapa bidang, terutama dalam klasifikasi hukuman TKP dan juga pengenalan format tanggal dan waktu dalam artikel berita online. Klasifikasi jenis kejahatan memiliki F1-score sebesar 92%, sedangkan skor F1 untuk Crime Location Extraction adalah 90,8%, dengan presisi 92,5%, dan recall 89,2%, dan untuk Crime Date Extraction, F1-score adalah 94,1%, dengan 100% presisi, dan 89% recall.