

ABSTRAK

Perkembangan teknologi digital yang pesat secara signifikan berpengaruh meningkatkan produktivitas dan kemudahan akses informasi dalam kehidupan sehari-hari, salah satunya penggunaan media sosial yang semakin meluas. Instagram merupakan salah satu *platform* yang banyak digunakan, dimana teks dalam *caption* memiliki informasi terkait lokasi yang dapat dimanfaatkan untuk analisis spasial. Penelitian ini bertujuan untuk mengidentifikasi dan mengklasifikasikan entitas lokasi dalam *caption Instagram* menggunakan algoritma *Support Vector Machine* (SVM) dengan pendekatan *Named Entity Recognition* (NER) *rule-based*. Metode yang digunakan meliputi ekstraksi fitur berbasis linguistik dengan konteks spasial eksplisit, lebelisasi data, pelatihan model, serta evaluasi kinerja model menggunakan matriks klasifikasi: akurasi, presisi, *recall* dan *f1-score*. Dataset terdiri dari 400 *captions* umumnya berbahasa Indonesia, namun terdapat unsur bahasa campuran seperti istilah asing atau bahasa daerah. Fokus utama penelitian diarahkan pada pengolahan dan pemahaman teks berbahasa Indonesia. *Dataset* dibagi menjadi 70% data *training* dan 30% data *testing*. Hasil pengujian menunjukkan bahwa model mendapatkan akurasi sebesar 90,83%, presisi 97,01%, *recall* 87,84% dan *f1-score* 92,90%. Evaluasi terhadap tiga *rules* NER (*exact match keyword*, pola preposisi, dan struktur deskriptif) menunjukkan bahwa pengenalan entitas berdasarkan gabungan seluruh *rules* memberikan *f1-score* tertinggi (89%), sementara *rule* individual terbaik adalah pola preposisi (74%). Nilai ini menunjukkan kinerja yang cukup baik dalam pengolahan *caption Instagram* yang variatif dan tidak terstruktur. Kombinasi metode SVM dan NER *rule-based* terbukti efektif dalam mengidentifikasi dan mengklasifikasi informasi spasial dalam dua kelas *Contain Location* dan *No Location*. Pendekatan ini berpotensi diterapkan pada sistem analisis spasial berbasis teks, seperti sistem rekomendasi lokasi, pemetaan geografis, dan pendukung keputusan berbasis lokasi.

Kata Kunci : *Instagram; Machine Learning; Named Entity Recognition; Support Vector Machine; Text Mining*

ABSTRACT

The rapid advancement of digital technology has significantly influenced productivity and facilitated access to information in daily life, particularly through the widespread use of social media. Instagram is one of the most popular platforms, where text in captions often contains location-related information that can be utilized for spatial analysis. This study aims to identify and classify location entities in Instagram captions using Support Vector Machine algorithm combine with rule-based Named Entity Recognition approach. The method involved linguistic feature extraction based on explicit spatial context, data labeling, model training, and performance evaluation using standard classification metrics: accuracy, precision, recall, and f1-score. Dataset consists of 400 captions primarily written in Indonesian, though some contain mixed-language elements such as foreign term or regional dialect. The dataset is divided into 70% training data ad 30% testing data. Experimental results show that model achieved an accuracy of 90,83%, precision of 97,01%, recall of 87,84%, and f1-score of 92,90%. Evaluation of three NER rules (exact match keyword, prepositional patterns, and descriptive structures) indicates that the combination of all rules yields the highest f1-score (89%), while the best-performing individual rule is the prepositioning pattern (74%). These results demonstrated strong performance in processing varied and unstructured Instagram captions. The combinations of SVM and NER rule-based prove effective in identifying and classifying spatial information into two classes Contains Location and No Location. This approach shows potential for implementation in text-based spatial analysis systems, such as location-based recommendation systems, geographic mapping, and location-based decision support systems.

Keywords: Instagram; Machine Learning; Named Entity Recognition; Support Vector Machine; Text Mining