ABSTRAK

SYARIFA NURFADILAH. Penerapan *Distant Supervision* Dan LSTM Berbasis *Doc2vec* Untuk Analisis Sentimen Aplikasi Jaki (Dibimbing oleh Ir. Ida, S.Kom., M.T. dan Fahrim Irhamna Rachman, S.Kom.,M.T.).

Penelitian ini mengusulkan sistem analisis sentimen otomatis ulasan aplikasi Jakarta Kini (JAKI) dengan mengintegrasikan Distant Supervision untuk pelabelan data, Doc2Vec untuk representasi teks, dan LSTM untuk pemodelan. Metode Distant Supervision rule-based berhasil melabeli otomatis ulasan ke dalam kelas positif, negatif, dan netral tanpa anotasi manual intensif. Representasi Doc2Vec Menghasilkan vektor dokumen kontekstual setelah praprosesan (tokenisasi, pembersihan, panjang 50). Model LSTM mencapai performa superior dengan F1-score tertimbang 96,35% pada skenario 70:30 (960 sampel test), akurasi 97,20%, precision dan recall 96,36%. Performa konsisten meningkat (90% → 94,53% → 96,35%) seiring bertambahnya data test, membuktikan robustness model pada dataset imbalanced. Pendekatan terintegrasi ini efisien, akurat, dan layak diimplementasikan untuk monitoring sentimen pengguna aplikasi layanan publik secara real-time.

Kata Kunci: Analisis Sentimen, Distant Supervision, Doc2Vec, LSTM, Jakarta Kini (JAKI), Word Embedding, Klasifikasi Multi-kelas, NLP Indonesia

ABSTRACT

SYARIFA NURFADILAH. *Implementation of Distant Supervision and Doc2Vec-Based LSTM for Sentiment Analysis of the JAKI Application (Supervised by Ir. Ida, S.Kom., M.T., and Fahrim Irhamna Rachman, S.Kom., M.T.).*

This research proposes an automated sentiment analysis system for Jakarta Kini (JAKI) app reviews integrating Distant Supervision for data labeling, Doc2Vec for text representation, and LSTM for modeling. Rule-based Distant Supervision successfully auto-labeled reviews into positive, negative, and neutral classes without intensive manual annotation. Doc2Vec representation generated contextual document vectors post-preprocessing (tokenization, cleaning, length 50). The LSTM model achieved superior performance with 96.35% weighted F1-score at 70:30 scenario (960 test samples), 97.20% accuracy, 96.36% precision and recall. Performance consistently improved (90% → 94.53% → 96.35%) with increasing test data, proving model robustness on imbalanced datasets. This integrated approach is efficient, accurate, and production-ready for real-time sentiment monitoring of public service app users.

Keywords: *Sentiment Analysis, Distant Supervision, Doc2Vec, LSTM, Jakarta Kini (JAKI), Word Embedding, Multi-class Classification, Indonesian NLP*