

INTISARI

MODEL DETEKSI KONTEN SPAM PADA MEDIA SOSIAL MELALUI PENDEKATAN PASANGAN POSTING-KOMENTAR MENGGUNAKAN LSTM-CNN BERBASIS *ATTENTION* DAN FITUR EMOJI

Oleh

ANTONIUS RACHMAT CHRISMANTO

20/468163/SPA/00729

Salah satu dampak negatif media sosial adalah banyaknya konten spam yang mengakibatkan kerancuan informasi. Komentar spam menjadi penyebab masalah tersebut yang perlu ditangani. Namun ketersediaan dataset bahasa Indonesia masih jarang, penelitian-penelitian sebelumnya melakukan deteksi komentar spam tanpa mempertimbangkan konteks posting yang dikomentari, dan mengabaikan emoji yang justru banyak ditemukan di media sosial. Penelitian ini mengusulkan deteksi komentar spam melalui model *deep learning* untuk bahasa Indonesia berbasis konteks dengan pendekatan baru menggunakan data pasangan posting-komentar dan fitur emoji.

Tahapan penelitian ini diawali dengan pengumpulan data pasangan posting-komentar dari media sosial Instagram dan menghasilkan kontribusi *dataset* SpamID-Pair menggunakan bahasa Indonesia. Tahapan berikutnya dilakukan *pre-processing* dan pembentukan vektor TF-IDF untuk metode *machine learning* dan vektor *embedding* dengan data *in-domain pretrained emoji aware* menggunakan Fasttext, GloVe, Word2Vec, dan Emoji2Vec, serta penggunaan *auxiliary features* untuk metode *deep learning*. Model *machine learning* dibangun dengan pendekatan baru menggunakan *ensemble machine learning* dari 14 metode *state-of-the-art* dan arsitektur *deep learning* dengan nama EiAP-BC (*Emoji Aware Inter-Attention Pair-BiLSTM CNN*) yang berbasis *inter-attention pair* untuk mendapatkan konteksnya.

Dari hasil penelitian, penggunaan fitur emoji, penggunaan pasangan data posting-komentar pada model *ensemble machine learning*, dan model *deep learning* EiAP-BC dapat diimplementasikan pada dataset SpamID-Pair dan terbukti meningkatkan performanya. Model EiAP-BC mempunyai rerata akurasi mencapai 88% dan 87% masing-masing untuk emoji-teks dan emoji-simbol. Meskipun belum mengalahkan, model yang diusulkan mampu bersaing dengan Google BERT standar dengan rerata akurasi mencapai 89%. Lebih dari itu model yang diusulkan memiliki keunggulan dalam hal komputasi yang lebih ringan, kesederhanaan arsitektur, pemrosesan yang lebih cepat, dan ukuran model lebih kecil. Selain itu, penelitian ini menghasilkan prototipe layanan web dan ekstensi browser untuk implementasi model yang dikembangkan.

Kata kunci: deteksi komentar spam, *ensemble ML*, EiAP-BC Model, media sosial

ABSTRACT

SPAM CONTENT DETECTION MODEL ON SOCIAL MEDIA USING POST-COMMENT PAIR APPROACH WITH ATTENTION BASED LSTM-CNN AND EMOJI FEATURE

By

ANTONIUS RACHMAT CHRISMANTO

20/468163/SPA/00729

Social media often suffers from spam content, which includes a lot of spam content that don't relate to the context of posts, making a misleading information. Spam comments are a problem that needs to be addressed. Unfortunately, detecting spam in Bahasa Indonesia remains challenging due to limited public datasets, the complexity of spam content, and the large number of emojis, which require context to detect it accurately. This research tries to detect spam comments with a deep learning model in Indonesian language based on the posts context using a new approach of post-comment pair data and emoji features.

The research methodology starts with collecting post-comment pairs from Instagram, resulting in a dataset named SpamID-Pair. Subsequently, data pre-processing is carried out, and features are generated in the form of TF-IDF vectors for machine learning. Additionally, embedding vectors are created using in-domain pre-trained emoji-aware models, including Fasttext, GloVe, Word2Vec, and Emoji2Vec, along with auxiliary features. These features are then processed by the ensemble machine learning of 14 state-of-the-art methods and a deep learning architecture known as the EiAP-BC (Emoji Aware Inter-Attention Pair-BiLSTM CNN) Model. This model focuses on understanding the relevance and context between comments and posts.

From experiments and testing results, the inclusion of emoji features, the post-comment pair approach in the ensemble machine learning model, and the EiAP-BC model enable context-based spam detection. The proposed model achieves an average accuracy of 88% and 87% for text-emoji and symbol-emoji, respectively on SpamID-Pair dataset. Notably, this performance competes favorably with Transformers BERT, which reaches 89% average accuracy. Importantly, the model has advantages like lower computational demands, simpler architecture, quicker processing, and a smaller model size. In addition, this research also produces web services and browser extensions prototype, making the model accessible to users.

Keywords: context based spam detection, ensemble machine learning, social media, EiAP-BC Model, LSTM-CNN, emoji.