

Analisa Long-Short Term Memory dan BERT Embeddings pada Klasifikasi Teks Data SMS Spam Berbahasa Indonesia = Analysis of Long-Short Term Memory and BERT Embeddings on Text Classification of SMS Spam Data in Indonesian

Zihan Nindia, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920545030&lokasi=lokal>

Abstrak

Pesatnya perkembangan teknologi informasi dan komunikasi telah membawa banyak perubahan dalam kehidupan manusia. Salah satu perkembangan yang paling signifikan adalah munculnya teknologi pesan singkat atau Short Message Service (SMS). Media SMS sering disalahgunakan sebagai media penipuan terhadap pengguna telepon. Penipuan sering terjadi dengan cara mengirimkan SMS secara masif dan acak hingga mencapai sepuluh ribu per hari kepada semua pengguna dan menjadi SMS spam bagi banyak orang. Klasifikasi teks menggunakan Long-Short Term Memory (LSTM) dan BERT Embeddings dilakukan untuk mengklasifikasi data SMS ke dalam dua kategori, yaitu spam dan non-spam. Data terdiri dari 5575 SMS yang telah diberi label. Dengan menggunakan metode LSTM + BERT, penelitian ini dapat mencapai nilai accuracy sebesar 97.85%. Metode ini menghasilkan hasil yang lebih baik dari ketiga model sebelumnya. Model LSTM + BERT menghasilkan nilai accuracy 0.65% lebih baik dari LSTM.

.....The rapid development of information and communication technology has brought many changes in human life. One of the most significant developments is the emergence of short message service (SMS) technology. SMS media is often misused as a medium for fraud against telephone users. Fraud often occurs by sending massive and random SMS up to ten thousand per day to all users and becomes SMS spam for many people. Text classification using Long-Short Term Memory (LSTM) and BERT Embeddings is performed to classify SMS data into two categories, namely spam and ham. The data consists of 5575 SMS that have been labeled. By using the LSTM + BERT method, this research can achieve an accuracy value of 97.85%. This method produces better results than the three previous models. The LSTM + BERT model produces an accuracy value of 0.65% better than LSTM.