

Implementasi Metode BERT-LSTM dalam Klasifikasi Asosiasi Single Nucleotide Polymorphisms dengan Fenotipe pada Penyakit Kronis Menggunakan Data Teks Ilmiah Bidang Kesehatan = Implementation of the BERT-LSTM Method for Classifying Associations Between Single Nucleotide Polymorphisms and Phenotypes in Chronic Diseases Using Scientific Text Data in the Health Domain

Inez Aurellia Rosyana, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920551535&lokasi=lokal>

Abstrak

Inovasi dalam bidang sains dan teknologi, salah satunya adalah perkembangan Artificial Intelligence, memiliki peranan yang besar bagi manusia untuk membantu menyelesaikan permasalahan yang ada. Tak terkecuali dengan bidang kesehatan atau medis yang merupakan salah satu sektor vital di kehidupan manusia. Precision medicine merupakan suatu konsep dalam perkembangan ilmu medis untuk memberikan diagnosis dan pengobatan yang presisi sesuai dengan karakteristik masing-masing pasien. Single Nucleotide Polymorphism (SNP) sebagai salah satu representasi biomarker dapat digunakan untuk melihat asosiasi genomik pasien dengan penyakit tertentu. Sejak tahun 2015, setiap tahunnya tercatat lebih dari 10.000 publikasi bidang kesehatan yang membahas mengenai SNP telah diterbitkan. Pengolahan data teks secara manual tentu memerlukan waktu, tenaga, dan biaya yang besar sehingga metode machine learning dimanfaatkan untuk dapat mengolahnya secara otomatis. Dalam mengolah data teks, teks perlu ditransformasi menjadi representasi numerik sebelum dilakukan proses machine learning selanjutnya. Penelitian ini menggunakan model BERT untuk mendapatkan representasi numerik dari data teks yang diberikan. Selanjutnya dalam membangun model klasifikasi, telah dikembangkan model deep learning yang salah satunya adalah neural network. Penelitian ini mengimplementasikan gabungan metode BERT untuk mendapatkan representasi data teks dan metode neural network LSTM untuk klasifikasi. Evaluasi performa model dilakukan dengan confusion matrix dengan metrik berupa accuracy, precision, recall, dan F1-score. Hasil dari penelitian ini didapatkan model gabungan BERT-LSTM memiliki performa yang lebih baik dibandingkan model LSTM yang menggunakan embedding dari library Pytorch dengan accuracy 86%, recall 82%, precision 86%, dan F1-score 83%.

..... Innovation in science and technology, one of which is the development of Artificial Intelligence, has a big role for humans to help solve existing problems. The health or medical field is no exception, which is one of the vital sectors in human life. Precision medicine is a concept in the development of medical science to provide precise diagnosis and treatment according to the characteristics of each patient. Single Nucleotide Polymorphism (SNP) as one of the biomarker representations can be used to see the genomic association of patients with certain diseases. Since 2015, more than 10,000 publications on SNPs have been published annually. Processing text data manually requires a lot of time, energy, and cost, so machine learning methods are utilized to process it automatically. In processing text data, the text needs to be transformed into a numerical representation before the next machine learning process is carried out. This research uses the BERT model to obtain a numerical representation of the given text data. Furthermore, in building a classification model, a deep learning model has been developed, one of which is a neural network. This research implements a combination of the BERT method to obtain text data representation and the LSTM

neural network method for classification. Performance evaluation is done with confusion matrix with metrics such as accuracy, precision, recall, and F1-score. The results of this study obtained that the combined BERT-LSTM model has better performance than the LSTM model using embedding from Pytorch library with 86% accuracy, 82% recall, 86% precision, and 83% F1-score.