

Akurasi transfer learning pada masalah analisis sentimen untuk tweets berbahasa indonesia = Accuracy of transfer learning on sentiment analysis problem for indonesian tweets

Kartika Syskya Wydya, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20455962&lokasi=lokal>

Abstrak

Analisis sentimen merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi. Pada penelitian ini, analisis sentimen diterapkan pada media sosial, yaitu Twitter. Pada dasarnya analisis sentimen merupakan masalah klasifikasi. Support Vector Machine SVM adalah salah satu metode machine learning untuk menyelesaikan masalah klasifikasi. Pada pendekatan SVM model dibangun dengan data dari domain yang sama. Namun, ketika terjadi perubahan domain, maka model machine learning harus dibangun kembali dari awal dengan menggunakan data pelatihan yang baru. Data pelatihan yang baru membutuhkan proses pelabelan yang dilakukan secara manual.

Dalam kasus ini, akan lebih efektif dan efisien jika dilakukan transfer learning agar dapat menggunakan data pelatihan dari domain yang sudah tersedia untuk menangani masalah klasifikasi pada domain yang berbeda.

Data pelatihan dari sebuah domain digunakan untuk melakukan klasifikasi pada domain yang berbeda.

Dalam penelitian masalah analisis sentimen untuk tweets berbahasa Indonesia ini, nilai akurasi transfer learning masih lebih rendah dari pada metode SVM tanpa transfer learning. Penggunaan fitur bi-gram dapat meningkatkan kinerja transfer learning.

Sentiment analysis is the process of understanding, extracting and processing textual data automatically to obtain information. In this experiment, sentiment analysis applied to social media, Twitter. Basically, sentiment analysis is a classification problem. Support Vector Machine SVM is one of machine learning method to solve two class classification problem. In the SVM approach the model is built with data from the same domain. However, when domain changes occur, the machine learning model must be rebuilt from scratch using new training data. New training data requires manual labeling process.

In this case, it would be more effective and efficient to transfer learning to use the training data from an already available domain to deal with classification problems on different domains. Training data from a domain will be used to classify on different domains. In the research problem of sentiment analysis for tweets in Bahasa, the value of transfer learning accuracy is still lower than the SVM method without transfer learning. Use of bi gram feature can improve the performance of transfer learning.