

Analisis penerapan algoritma support vector machine pada sistem deteksi plagiarisme berbasis latent semantic analysis = Analysis of support vector machine algorithm implementation in plagiarism detection system based on latent semantic analysis

Dian Rismawati, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20431447&lokasi=lokal>

Abstrak

Departemen Teknik Elektro Universitas Indonesia telah mengembangkan suatu sistem berbasis Latent Semantic Analysis (LSA) untuk mendeteksi plagiarisme pada karya tulis berbahasa Indonesia dan Inggris. Data keluaran sistem deteksi plagiarisme berbasis LSA adalah nilai frobenius norm, slice, dan pad. Pada skripsi ini akan menjelaskan serta memberikan analisis pada pengembangan sistem deteksi plagiarisme yang telah ada yaitu dengan menerapkan algoritma Support Vector Machine (SVM).

Support Vector Machine (SVM) adalah suatu Learning Algoritm yang bertujuan untuk menemukan suatu hipotesis berupa bidang pemisah (hyperplan) terbaik dari sekumpulan data yang dapat dipisahkan secara linear maupun tidak linear. SVM akan memisahkan data hasil keluaran sistem deteksi plagiat berbasis LSA menjadi dua kelas yaitu "plagiat" dan "tidak plagiat" dengan menggunakan 2 metode yaitu kombinasi data input dan kombinasi data output dengan metode AND. Beberapa modifikasi terhadap input program dilakukan diantaranya memvariasikan parameter-parameter pembelajaran dan memvariasikan data hasil keluaran program deteksi plagiarisme berbasis LSA.

Hasil dari analisis serta pengujian yang telah dilakukan yaitu jika menggunakan parameter serta kombinasi data yang tepat, SVM mampu untuk meningkatkan akurasi sistem dari sistem yang menggunakan metode Learning Vector Quantization (LVQ) pada penelitian sebelumnya hingga menghasilkan akurasi sebesar 63,15% hal ini dilihat jika mempertimbangkan keseimbangan terhadap aspek presisi dan relevansi program sedangkan jika dilihat melalui presentase jumlah data yang berhasil diklasifikasikan dengan tepat, SVM mampu menghasilkan akurasi sebesar 97,04%.

Department of Electrical Engineering, University of Indonesia has developed a system based on Latent Semantic Analysis (LSA) to detect plagiarism between two paper written in different languages, which are Indonesian and English. The output data of plagiarism detection system are frobenius norm, slice, and pad. This thesis will explain and provide analysis of the development of plagiarism detection system that already exist by applying Support Vector Machine (SVM) algorithm.

Support Vector Machine (SVM) is a Learning Algorithm that aims to find a best hypothetical form called hyperplan to separated a set of data that can be separated linearly and nonlinearly. SVM will separate output data of plagiarism detection system into two classes, "plagiat" class and "tidak plagiat" class by using two methods: combination of input data method and output data combined with AND method. Some modifications to input program are made, such as variating the parameters of learning and variating the output data of plagiarism detection program.

The results of analysis and test that has been done are: if the system use correct parameters and correct combinations of the data, SVM is able to improve accuracy of the system from the last research that using Learning Vector Quantization (LVQ). The accuracy of SVM is 63,15% if considering the balance of precision and relevance of the program, while when viewed through a percentage of the amount of data that

appropriately classified, the accuracy of SVM is 97.04%.</i>