

Abstrak

Pada *supervised machine learning*, data latih yang telah diberi label yang benar merupakan suatu hal yang menjadi prasyarat. Namun, dalam banyak pengaplikasiannya, tugas memberi label tidak bisa dilakukan secara otomatis, tapi melibatkan keputusan manusia dan oleh karena itu membutuhkan waktu yang banyak dan mahal.

Pada tugas akhir ini, *active learning* diimplementasikan pada *support vector machine* dan diteliti apa faktor yang mempengaruhi jumlah data latih yang diberi label dan tingkat akurasi sistem, dan bagaimana pengaruhnya. Selain itu juga dibandingkan metode pemilihan inisial data dan *next data*, yaitu metode random dan metode *dissimilarity*. Data yang digunakan dalam tugas akhir ini adalah *Winsconsin Breast Cancer Diagnosis* dan *Hill-Valley* dari *UCI Repository*. Tujuan utama dari *active learning* adalah memilih data yang penting atau berpengaruh pada sistem, sehingga bisa mengurangi jumlah data yang perlu diberi label.

Hasil penelitian adalah *active learning* mampu mengurangi jumlah data yang harus diberi label sampai 82.5% tanpa terjadi penurunan akurasi sistem yang *significant*.

Kata kunci : *Active Learning, Support Vector Machine*, klasifikasi, label data, pengurangan.