

Abstrak

Menemukan informasi dari dokumen yang sangat banyak dan bertambah setiap harinya adalah tugas yang sulit dan memakan waktu. Untuk mendapatkan dokumen yang diinginkan diantara sekian banyak dokumen yang ada dilakukan dengan cara membaca setiap dokumen satu per satu. Hal ini dapat diatasi dengan pemetaan dokumen dalam *cluster* yang berbeda (*clustering*) berdasarkan informasi yang berada didalamnya.

Clustering dilakukan untuk mengelompokkan data ke dalam satu klaster dimana objek pada klaster yang sama memiliki kemiripan yang sangat besar. Namun terdapat permasalahan dalam proses klasterisasi yaitu adanya data berdimensi tinggi. Untuk itu diperlukan algoritma yang mampu menangani hal ini. *Self Organizing Maps* (SOM) adalah algoritma yang tepat untuk klasterisasi data dengan jumlah dimensi yang tinggi. SOM merupakan algoritma jaringan saraf tiruan (JST) yang bekerja dengan cara memperbaharui nilai bobot beratnya sehingga cocok dengan nilai vektor input. Dengan cara kerja ini, jaringan yang berisi neuron-neuron akan menyusun dirinya sendiri berdasarkan input nilai tertentu dalam suatu *cluster*.

Fokus dari tugas akhir ini adalah implementasi algoritma SOM pada klasterisasi dokumen berita berbahasa Indonesia. Pengujian dilakukan dengan menganalisis pengaruh jumlah dimensi dan parameter input seperti jumlah epoch dan learning rate terhadap performansi dan akurasi sistem. Di akhir pengujian dapat di buktikna bahwa SOM mampu menangani data berdimensi tinggi dengan baik. Perhitungan performansi hasil klasterisasi dilakukan berdasarkan *precision* dan *entropy*.

Kata kunci: *Clustering, Unsupervised learning, Self Organizing Maps.*