

---

**Abstrak**

Seiring berkembangnya zaman perkembangan teknologi juga memiliki peningkatan. Persebaran informasi sangat cepat terjadi pada media sosial khususnya twitter. Dalam twitter tidak semua berita yang beredar belum tentu merupakan informasi yang benar. Banyak sekali informasi yang disebarluaskan merupakan berita hoax yang disebarluaskan oleh oknum yang tidak bertanggung jawab. Dalam tugas akhir ini, penulis akan membangun sebuah sistem yang dapat menentukan jumlah data train yang optimal dalam proses klasifikasi berita hoax. Pada penelitian ini penulis akan memakai algoritma support vector machine dan word2vec untuk melakukan klasifikasi berita hoax dan non hoax pada sistem yang akan dibuat. Pada penelitian ini dilakukan percobaan sebanyak 5 kali dengan jumlah data train yang digunakan sebanyak 5000, 10000, 15000, 20000, 25000. Data train sebanyak 5000 menghasilkan akurasi sebesar 77,28%, data train sebanyak 10000 menghasilkan akurasi sebesar 79,68%, data train sebanyak 15000 menghasilkan akurasi sebesar 79,892%, data train sebanyak 20000 menghasilkan akurasi sebesar 80,416%, dan data train sebanyak 25000 menghasilkan akurasi sebesar 81,184%. Performa akurasi terbaik didapat percobaan dengan jumlah data train sebanyak 25000 mendapatkan performa akurasi terbaik sebesar 81,184% dengan menggunakan kombinasi unigram dan seleksi token full token.

**Kata kunci :** Hoax, Klasifikasi, Support Vector Machine, Word2Vec, Twitter