

Abstrak

Saat ini penyebaran informasi *Hoax* sangat marak terjadi di berbagai media. Salah satu media nya yaitu berita. Banyak penelitian telah dilakukan dengan metode pengelompokan berita menggunakan *K-Means*, akan tetapi penelitian yang ada lebih banyak menggunakan data berita yang asli. Pada penelitian ini, terdapat algoritma *K-Means* yang digunakan. Penelitian ini menggunakan dataset 1.231 *record* yang berisi *Hoax* dan digunakan 16 kelompok yang terdiri dari *advice fee, bogus warnings, death, facebook scams, fake news, faux images, health, general, legend, malware, misleading, lottery, scams, special featurers*. Lalu dataset akan dilakukan proses pre-processing terlebih dahulu. Untuk tahap preprocessing dibagi menjadi beberapa bagian, yakni *clear url, clear symbol, tokenisasi, stop words, normalisasi, garbage*. Lalu hasil dari *preprocessing* diolah pada proses pembobotan kata. Setelah melewati tahap pembobotan kata, maka dapat dikelompokkan dengan menggunakan metode dari *k-means clustering*. Pada penelitian ini dilakukan dengan memasukkan jumlah cluster yang bervariasi. Dari hasil validasi dengan memasukkan nilai cluster yang berbeda telah didapatkan nilai cluster yang terbaik yaitu $K=16$ dengan nilai $SSE(Sum Square Error)$ yang dihasilkan 1058.0332293029496.

Kata Kunci : *E-Mail, Hoax, Clustering, Algoritma K-Means, SSE(Sum Square Error)*.