

1. Pendahuluan

Latar Belakang

Pada era digital seperti sekarang, penyebaran informasi melalui internet sudah sangat bebas dan bisa diakses dimanapun selama terdapat koneksi internet. Imbasnya, akhir-akhir ini di Indonesia banyak sekali berita yang berasal dari situs berita *hoax* disebar melalui media sosial. Hal ini sangat memprihatinkan, terutama dengan makin maraknya situs-situs berita *hoax* bermunculan. Hal tersebut terbukti berdasarkan informasi dari Kementerian Komunikasi dan Informatika bahwa pada tahun 2016 ada 800 ribu situs di Indonesia yang terindikasi sebagai penyebar berita palsu dan ujaran kebencian [1]. Selain itu, ada dua kelompok penyebar situs berita *hoax* berskala besar yang tertangkap polisi [2]. Berdasarkan survey yang dilakukan kepada masyarakat, 98.7% responden setuju bahwa berita *hoax* mengganggu kerukunan karena mayoritas berita *hoax* membahas tentang sosial politik dan SARA. Selain itu, 34.9% penyebaran berita *hoax* berasal dari situs web [3]. Oleh karena itu perlu ada sistem yang dapat menyaring dan mencegah penyebaran berita *hoax* agar tidak lagi banyak terjadi orang yang pemikirannya tersesat dengan adanya berita *hoax*.

Pada penelitian ini, akan ditawarkan suatu sistem yang dapat mendeteksi apakah suatu berita yang berasal dari situs web termasuk kedalam berita *hoax* atau bukan. Dalam mengklasifikasi berita, terdapat tiga tahapan, yaitu *text pre-processing*, *feature extraction* menggunakan TF-IDF, dan klasifikasi menggunakan *Support Vector Machine* (SVM). Metode TF-IDF dan SVM dipilih karena terbukti dapat menghasilkan akurasi yang tinggi [4].

Topik dan Batasannya

Pada penelitian ini sistem pendeteksi berita *hoax* dibangun untuk menyelesaikan persoalan mengenai maraknya penyebaran berita *hoax* melalui internet. Sistem nya terdiri dari tiga bagian, yaitu *text preprocessing*, *feature extraction*, klasifikasi. Sebanyak 205 data berita (100 berita *hoax* dan 105 berita bukan *hoax*) akan digunakan untuk proses pelatihan dan pengujian menggunakan metode 10 *fold-cross validation*. Pengujian dilakukan dengan mempertimbangkan penggunaan proses *stemming* dan tanpa *stemming* untuk melihat proses *text preprocessing* mana yang menghasilkan akurasi paling baik pada penelitian ini.

Dalam mengklasifikasi berita berbahasa Indonesia, *input* teks berita akan melalui proses *text preprocessing* sehingga menjadi bagian-bagian kata. Setelah itu, dilakukan *feature extraction* pada setiap kata menggunakan TF-IDF yang menghasilkan suatu nilai. Nilai TF-IDF tersebut menjadi data yang nantinya diolah oleh *classifier* SVM dan menghasilkan *output* kelas *hoax* atau bukan *hoax*. Berita yang dapat diklasifikasi hanya berita yang berasal dari situs berita dan berbahasa Indonesia. Karena pada proses *text preprocessing* menggunakan metode khusus untuk bahasa Indonesia, mulai dari kata dasar, *stopword*, hingga *stemming*.

Tujuan

Penelitian ini bertujuan untuk membuat sistem pendeteksi berita *hoax* berbahasa Indonesia dan membandingkan proses pada *text preprocessing* dengan menggunakan *stemming* dan tanpa *stemming*. Kemudian, dilakukan percobaan dengan menggunakan parameter *classifier* SVM yang berbeda-beda. Perbandingan proses pada *text preprocessing* dan Percobaan parameter SVM yang berbeda-beda dilakukan untuk memastikan konfigurasi terbaik untuk klasifikasi berita *hoax*.

Organisasi Tulisan

Setelah pendahuluan yang sudah dijelaskan di atas. Bagian-bagian selanjutnya pada penelitian ini berisi Studi Terkait yang menjelaskan mengenai penelitian-penelitian sebelumnya, Sistem yang dibangun menjelaskan alur penelitian dan alur perancangan sistem, Evaluasi yang menjelaskan hasil dan analisa berdasarkan pengujian, dan Kesimpulan dari keseluruhan penelitian.