

1. PENDAHULUAN

1.1. Latar Belakang

Dalam perkembangan teknologi yang pesat di era modern, sosial media sangatlah populer di kalangan masyarakat. Berdasarkan survey yang diambil dari tahun 2005 – 2015 di AS, penggunaan sosial media telah meningkat secara drastis [[1]]. Di Indonesia sendiri pada tahun 2020 ini pengguna internet sudah mencapai 175,4 juta pada bulan Januari, sedangkan pengguna media sosial sudah mencapai 160,0 juta dengan peningkatan sebanyak 12 juta (8,1%) dari tahun 2019 sampai Januari 2020[2]. Popularitas media sosial memang tidak bisa dibendung yang artinya tidak ada yang bisa menghentikan menyebarkan fakta atau berita palsu[3]. Di sisi lain perkembangan media sosial ibarat pisau dua sisi semua orang menggunakan media sosial untuk bekerja, belajar, berkomunikasi, dan lain – lainnya.

Namun, media sosial kini sudah sering terganggu karena isu – isu berita *hoax* yang mengakibatkan pengguna media sosial saat ini sulit untuk membedakan informasi berita asli atau berita palsu (*Hoax*) terutama pada informasi yang tersebar luas di twitter[4]. Oleh karena itu diperlukan suatu metode atau algoritma yang mampu mendeteksi berita palsu. Pada tren untuk klasifikasi dataset dijelaskan oleh para ilmuwan dengan menggunakan metode *machine learning*, algoritma *machine learning* merupakan salah satu metode paling efektif dalam mengklasifikasikan atau memfilter data informasi *hoax*.

Terkait riset – riset yang sudah dilakukan sebelumnya seperti Darmawan A “Hoax News Analysis for the Indonesian National Capital Relocation Public Policy with the Support Vector Machine and Random Forest Algorithms” memperoleh akurasi SVM sebesar 95,24% dibandingkan dengan *Random Forest* sebesar 86,90% [5]. Kemudian selanjutnya penelitian oleh Sagita R ,dll yang berjudul “Klasifikasi berita clickbait menggunakan *K-Nearest Neighbor*” berhasil memperoleh hasil terbaik pada $k=11$ menggunakan scenario pembagian data dengan jumlah data 800 dan data uji 200 menghasilkan akurasi sebesar 71%, presisi sebesar 72%, dan *recall* sebesar 71% [6]. Dan yang terakhir penelitian oleh Ramadhan N, dll yang berjudul “Deteksi Berita Palsu Menggunakan Metode

Random Forest dan *Logistic Regression*” memperoleh akurasi untuk model *Random Forest* sebesar 84% sedangkan untuk model *Logistic Regression* sebesar 77%[7].

Sehingga pada penelitian algoritma machine learning yang akan digunakan penulis adalah algoritma *Random Forest*. Kelebihan *Random Forest* adalah metode *averaging ensemble* yang dapat digunakan untuk masalah proses klasifikasi dan juga mengatasi *overfitting*. Metode *Random Forest* untuk mengklasifikasikan teks ini telah dibandingkan dengan metode lain salah satunya yaitu *Adaboost*. Penelitian tersebut menunjukkan bahwa mode *Random Forest* memiliki akurasi sebesar 93,58% dan f1-score 89,65% lebih tinggi dibanding *adaboost* dalam mendeteksi berita palsu (*hoax*)[8].

1.2. Perumusan Masalah

Rumusan masalah yang akan diangkat pada penelitian ini adalah :

1. Bagaimana performansi dari algoritma klasifikasi *Random Forest* dalam mendeteksi berita palsu (*hoax*)?
2. Bagaimana pengaruh ekstraksi fitur TF-IDF pada performansi algoritma klasifikasi *Random Forest*-Apriori dalam mendeteksi berita palsu (*hoax*)?

1.3. Tujuan

Berikut adalah tujuan yang ingin dicapai dari penulisan proposal :

1. Mengetahui performansi dari algoritma klasifikasi *Random Forest* dalam mendeteksi berita palsu (*hoax*).
2. Mengetahui pengaruh ekstraksi fitur TF-IDF pada performansi algoritma klasifikasi *Random Forest*-Apriori dalam mendeteksi berita palsu (*hoax*).

1.4. Rencana Kegiatan

Rencana kegiatan yang akan dilakukan pada penelitian ini adalah sebagai berikut :

- a. Studi literature

Studi literatur dilakukan untuk observasi mengenai permasalahan pada penelitian ini berdasarkan buku, *paper*, atau jurnal yang sudah melakukan

penelitian sejenis sebelumnya. Pencarian berbagai sumber informasi dan metode yang sesuai dengan domain maupun diluar domain permasalahan akan digunakan dalam menunjang penelitian.

b. Perancangan Sistem

Tahap pertama dalam perancangan system adalah pre-processing data, lalu dilanjut dengan ekstraksi fitur menggunakan TF-IDF kemudian diklasifikasikan menggunakan Random Forest.

c. Penyusunan Laporan

Hasil dari seluruh Tugas Akhir ini didokumentasikan dalam bentuk laporan.

1.5. Jadwal Kegiatan

Jadwal pelaksanaan dibuat berdasarkan rencana kegiatan. Bar-chart bisa dibuat per bulan atau per minggu. Contoh bar-chart:

Table 1. Jadwal Kegiatan

Kegiatan	Bulan					
	1	2	3	4	5	6
Studi Literatur						
Perancangan Sistem						
Pengumpulan data						
Crawling, Pelabelan, Pre-processing						
Implementasi Algoritma random forest						
Pengujian Sistem						
Penyusunan laporan						

*Keterangan: shading warna *grayscale*