

BAB I

PENDAHULUAN

1.1 Latar Belakang

Berita adalah informasi yang dapat menarik minat pembaca sekaligus informasi terhangat yang disampaikan melalui media sosial, surat kabar, media online, dan lain-lain. Hal ini disebabkan pesatnya perkembangan teknologi informasi, serta penggunaan internet yang bebas mendapatkan informasi dari media sosial atau sejenisnya. Banyak media massa yang memanfaatkannya sebagai media penyebaran berita. Namun pada kenyataannya banyak sekali berita *hoax* yaitu operasi berita yang dilakukan dengan di sengaja dan dimaksudkan untuk memberikan pemahaman palsu. Singkatnya *hoax* adalah berita yang mengandung informasi palsu atau tidak benar yang tidak dapat dipertanggungjawabkan oleh orang tertentu dan bertujuan untuk meyakinkan masyarakat (awam) agar tidak tertipu dan percaya dengan berita *hoax* tersebut. Dengan banyaknya sebaran berita yang sangat luas, dan akan sangat sulit untuk memberikan informasi mana berita yang *valid* dan mana berita yang *hoax* [1].

Adanya berita untuk memberikan informasi terbaru yang sudah terjadi atau yang akan terjadi. Adapun tujuannya dimaksudkan untuk mempengaruhi masyarakat luas karena itulah berita yang disajikan harus berdasarkan *valid* karena berpengaruh kepada masyarakat. Perkembangan teknologi yang sangat cepat, berita *hoax* mudah tersebar melalui internet, sehingga masyarakat awam mudah terkecoh dan sulit menentukan suatu berita *hoax* atau benar adanya. Berdasarkan permasalahan tersebut maka dibutuhkan sistem klasifikasi berita *hoax* [2].

Penelitian sebelumnya yang dilakukan oleh Andre Rino Prasetyo menggunakan Metode *Modified K-Nearest Neighbor* tentang berita *hoax* dengan data berita kesehatan dari portal berita dan jejaring sosial yang diperoleh secara manual, dimana pakar ahli memberikan label total 51 berita, tim *hoax buster* memberikan label 67 dan sisanya diberi label oleh situs portal berita 52 berita, dengan input berupa teks ini dapat digunakan pada sistem klasifikasi berita kemudian diolah menggunakan beberapa tahapan proses, peratama yaitu tahapan

preprocessing, kemudian tahapan pembobotan kata TF-IDF, selanjutnya tahapan *cossine similarity*, dan menghitung nilai validitas, yang terakhir yaitu tahapan perhitungan *weight voting* dimana pada tahap ini adalah mengurutkan dan memutuskan pengelompokan terbesar berdasarkan nilai *k-values*. Dengan perhitungan tersebut memanfaatkan metode *Modified K-Nearest Neighbor* memperoleh akurasi sebesar 80%. Hasil pengujian *k-values* yang bernilai 4 memperoleh akurasi tertinggi 75%, dengan nilai presisi 83%, *recall* 75% dan *f-measure* 79% [3]. Perbedaan pada penelitian ini yaitu ada di jumlah data, dimana pada penelitian saya data menggunakan sebanyak 600 data atau dokumen.

Sistem klasifikasi berita *hoax* terdiri dari beberapa tahapan yaitu tahapan *preprocessing*, kemudian pembobotan kata TF-IDF, dan *cossine similarity*, kemudian proses klasifikasi dengan metode *Modified K-Nearest Neighbor*. Dengan input berupa data teks berita kemudian diolah untuk menghasilkan output berupa *hoax* atau hasil yang *valid*. Data yang digunakan dihasilkan dari berbagai kutipan berita di Mendeley.

Berdasarkan penjelasan di atas berdasarkan penelitian [3] terbukti bahwa metode *Modified K-Nearest Neighbor* memiliki tingkat akurasi yang sangat baik. Metode ini dipilih karena sangat tepat digunakan untuk melihat data yang terindikasi *hoax* atau *valid*. Dengan adanya sistem klasifikasi berita *hoax* dengan metode *Modified K-Nearest Neighbor* ini dapat membantu masyarakat untuk tidak mudah mempercayai berita dengan sumber yang tidak jelas.

1.2 Perumusan Masalah

1.2.1 Permasalahan

Banyak penelitian sebelumnya membahas tentang berita *hoax* tapi belum memberikan akurasi yang bagus sehingga dibutuhkan suatu metode untuk meninjau kembali dengan menambahkan lebih banyak data pada prosesnya untuk mendapatkan akurasi lebih baik dengan menggunakan metode *Modified K-Nearest Neighbor* (KNN) yang merupakan perbaikan dari metode *K-Nearest Neighbor* (KNN).



1.2.2 Metode Usulan

Pemanfaatan metode *Modified K-Nearest Neighbor* (KNN) untuk klasifikasi berita *hoax* yang merupakan perbaikan dari metode *K-Nearest Neighbor* dengan penambahan perhitungan *validity* dan *weight voting* untuk meningkatkan akurasi yang cenderung menghasilkan akurasi yang kurang baik.

1.2.3 Pertanyaan Penelitian

Berapa tingkat akurasi yang diperoleh dalam klasifikasi berita *hoax* dengan menggunakan metode *Modified K-Nearest Neighbor* (KNN).

1.3 Tujuan dan Manfaat

1.3.1 Tujuan Penelitian

Untuk mengetahui tingkat akurasi yang diperoleh berdasarkan nilai validitas data latih dan nilai validitas data uji yang telah ditentukan dalam proses klasifikasi berita *hoax* dengan menggunakan metode *Modified K-Nearest Neighbor* (KNN) dengan menambahkan lebih banyak data dari pada penelitian sebelumnya.

1.3.2 Manfaat

Manfaat yang diharapkan dalam penelitian ini adalah untuk mendapatkan akurasi yang lebih baik dari penelitian sebelumnya.

1.4 Batasan-batasan

Batasan-batasan masalah pada penelitian ini adalah menggunakan data berita umum yang diambil dari situs *Mendeley* tentang *Indonesian Hoax News Detection Dataset* periode 2018 dengan jumlah data 600.

1.5 Sistematika Penulisan

Adapun sistematika penulisan adalah sebagai berikut :

1. BAB I PENDAHULUAN

Pada bab I ini berisikan Latar Belakang, Rumusan Masalah, Tujuan dan Manfaat, Batasan-batasan, dan Sistematika Penulisan.

2. BAB II KAJIAN PUSTAKA

Pada bab II ini berisikan landasan teori yang digunakan yaitu Klasifikasi teks, Berita, *Hoax*, *Modified K-Nearest Neighbor (KNN)*, *Cosine Similarity*, *Weight Voting*, dan Penelitian Sebelumnya.

3. BAB III METODE USULAN

Pada bab III ini berisikan tentang Flowchart Sistem, Arsitektur M-KNN, Dataset yang digunakan, dan Tahapan Penelitian.

4. BAB IV HASIL DAN PEMBAHASAN

Pada bab IV ini berisikan tentang Lingkungan Uji Coba, Pembahasan dan Hasil dari Perhitungan Program pada penelitian.

5. BAB V PENUTUP

Pada bab V ini berisikan tentang Kesimpulan dan Saran pada penelitian.