

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi telah menjadikan dunia tanpa batas dan membawa perubahan sosial yang cepat. Teknologi informasi saat ini, selain membantu meningkatkan kesejahteraan, kemajuan dan peradaban umat manusia, juga menjadi lahan subur bagi terjadinya perbuatan melawan hukum. Contoh pelanggaran hukum yang sangat umum terjadi saat ini melalui penggunaan teknologi informasi adalah penyebaran berita palsu di jejaring sosial[1].

Hoax merujuk kepada informasi yang mencakup elemen yang tidak dapat dipercaya dan juga *hoax* memiliki informasi yang tidak sesuai dengan suatu kejadian yang terjadi pada kenyataannya, dalam konteks *hoax* itu sendiri memiliki tanda-tanda bahwa berita tersebut pasti berasal dari sumber yang tidak jelas keberadaannya jika informasi *hoax* tersebut masuk ke dalam kalangan masyarakat tentunya bisa berdampak negatif dan juga dengan adanya *hoax* bisa merusak kepercayaan publik, dan konsekuensi lainnya[2].

Salah satu upaya pencegahan penyebaran berita *hoax* adalah dengan melakukan identifikasi untuk memastikan apakah informasi berita tersebut yang dibaca tersebut benar atau bukan. Hal ini bertujuan untuk memberikan penjelasan kepada pembaca. Kini telah dilakukan sebagian penelitian mengenai pendeteksian *hoax* dengan memanfaatkan model pembelajaran mesin atau kecerdasan buatan dengan kehadiran penelitian ini tentunya sangat bermanfaat bagi pembaca berita dapat menjauhi dari berita *hoax*[3]. Sebuah penelitian tentang pendeteksian berita palsu pernah dilakukan oleh Candra Surya Sriyano. Mereka melakukan pengklasifikasian berita *hoax* menggunakan metode *Naive Bayes*, yang berhasil mencapai tingkat akurasi sebesar 72% [4].

Dalam penelitian sebelumnya, klasifikasi berita palsu dilakukan menggunakan metode *K-Nearest Neighbor* (KNN) dan *Naive Bayes*. Hasil akurasi yang diperoleh dari metode KNN adalah sekitar 68%, sementara metode *Naive Bayes* menghasilkan tingkat akurasi sekitar 66%.[5]. *K-Nearest Neighbor* adalah sebuah algoritma pembelajaran berbasis pengawasan (*supervised learning*) yang memungkinkan klasifikasi pada sekumpulan data kemudian penelitian lainnya dengan melakukan klasifikasi berita *hoax*, pada penelitian tersebut menggunakan metode *K-Nearest Neighbor* [6].

Hasil akurasi yang dihasilkan adalah 48% dengan menggunakan akurasi model dengan menggunakan *k-fold cross validation*. Hal ini karena beberapa kelas memiliki yang tidak merata, sehingga hasil yang didapat kurang maksimal[6]. Salah satu langkah yang dapat diambil adalah melalui modifikasi metode *K-Nearest Neighbor* untuk mengatasi ketidakmerataan data teks dalam pendeteksian berita *hoax* yaitu dengan metode NWKNN, dimana dengan metode ini memakai dasar pembobotan yang dimaksud pembobotan adalah bobot akan lebih rendah diberikan kepada tetangga-tetangga yang berasal dari kelas mayoritas, dan sebaliknya, bobot yang lebih tinggi diberikan kepada tetangga-tetangga yang berasal dari kelas minoritas[7].

Penelitian yang dilakukan oleh Afrizal Rivaldi, dkk mengenai NWKNN, terbukti bahwa sebuah model NWKNN menghasilkan akurasi lebih tinggi dibandingkan dengan KNN dengan hasil 95% memberikan hasil akurasi 5%-15% lebih baik dari pada metode KNN[8]. Kemudian pada penelitian lain yang dilakukan oleh Martha memberikan hasil bahwa metode NWKNN memiliki hasil lebih tinggi dibanding dengan KNN, dengan hasil 86 % untuk NWKNN dan 82% untuk KNN[9].

Oleh karena itu, pada penelitian ini penulis membahas apa pengaruh dari peformansi sistem yang diterapkan menggunakan algoritma Neighbor Weighted K-Nearest Neighbor (NWKNN).

1.2 Rumusan Masalah

Berdasarkan latar belakang penelitian tersebut, maka perumusan masalah dalam penelitian ini yaitu bagaimana performansi metode *Neighbor Weighted K-Nearest Neighbor* (NWKNN) dalam melakukan deteksi *hoax*.

1.3 Tujuan Penelitian

Tujuan penelitian ini untuk mengetahui performansi yang dihasilkan dari mengklasifikasikan berita *hoax* dengan metode metode *Neighbor Weighted K-Nearest Neighbor* (NWKNN) untuk mengatasi masalah data teks yang tidak terdistribusi secara rata sehingga mempengaruhi akurasi.

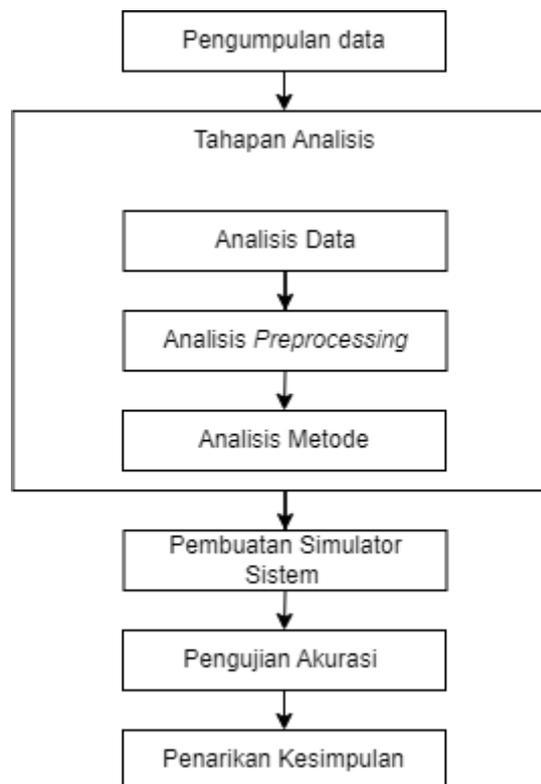
1.4 Batasan Masalah

Adapun batasan masalah pada penelitian yang akan dilakukan antara lain:

1. Sumber dataset diambil dari situs *Kaggle* dengan kata kunci “Indonesia False News(Hoax) Dataset”.
2. Dataset yang diambil berupa data bahasa Indonesia.
3. Klasifikasi berita hoax terdiri atas kelas hoax dan non hoax.
4. Format file yang digunakan pada data latih dan uji adalah csv.
5. Kata yang diinputkan tidak menggunakan emoticon.

1.5 Metodologi Penelitian

Metode penelitian yang dipakai untuk penelitian ini adalah penulis akan menggunakan metode kualitatif. Penelitian kualitatif adalah suatu metode penelitian yang berlandaskan filsafat positivisme, digunakan untuk mempelajari populasi atau sampel tertentu, mengumpulkan data dengan menggunakan alat penelitian, menganalisis data secara kuantitatif atau statistik, bertujuan untuk menguji hipotesis yang telah ditentukan[10].



Gambar 1.1 Metodologi Penelitian

1.5.1 Pengumpulan Data

Pengumpulan data yang digunakan pada penelitian ini adalah sebagai berikut:

1. Studi Literatur, yaitu teknik pengumpulan data dengan cara mengumpulkan literatur, jurnal, paper, dan buku yang berkaitan dengan analisis berita hoax.
2. Pengumpulan Dataset, dataset yang digunakan pada penelitian ini berupa validasi berita di Indonesia yang diambil dari situs *Kaggle* dengan cara manual.

1.5.2 Tahapan Analisis

Berdasarkan pengkajian dan evaluasi setelah membaca beberapa sumber seperti jurnal, dan buku – buku, Tahapan analisis pada penelitian ini adalah:

1. Analisis Data, yaitu menganalisis pola kalimat dari data yang didapat, data yang dianalisis adalah kalimat dari validasi berita mengenai berita di Indonesia.
2. Analisis Preprocessing, Pada tahap ini, kalimat narasi yang terkumpul akan diproses sehingga data yang didapat menjadi lebih terstruktur dan mudah untuk diolah. Langkah-langkah preprocessing yang dilakukan terdiri dari *case folding*, *cleaning data*, *tokenization*, *stopword*, *stemming*.
3. Analisis Metode, Pada tahap ini akan dilakukan analisis klasifikasi menggunakan metode *Neighbor Weighted K-Nearest Neighbor* (NWKNN) meliputi pembobotan kata dengan metode TF-IDF, penentuan jarak ketetanggaan dengan metode Euclidean Distance, dan perhitungan skor.

1.5.3 Pembuatan Simulator Sistem

Dalam tahapan ini akan menjelaskan tentang implementasi pembuatan simulator sistem dengan metode yang dibuat dengan menggunakan metode *Neighbor Weighted K-Nearest Neighbor* (NWKNN) untuk mengklasifikasi berita hoax di Indonesia. Simulator sistem yang akan digunakan adalah bahasa pemrograman Python.

1.5.4 Pengujian

Dalam tahapan ini akan dilakukan pengujian akurasi dari metode algoritma NWKNN yang akan diimplementasikan ke sistem deteksi berita hoax terhadap validasi berita hoax di Indonesia pada situs *Kaggle* dengan menampilkan akurasi pada setiap k.

1.5.5 Penarikan Kesimpulan

Pada tahap ini akan dilakukan penarikan kesimpulan yang berdasarkan hasil dari tahap pengujian dalam penerapan metode NWKNN pada kasus deteksi berita hoax.

1.5.6 Sistematika Penulisan

Sistematika penulisan disusun untuk memberikan gambaran umum tentang penulisan skripsi yang akan dilakukan. Sistematika penulisan tugas akhir ini adalah sebagai berikut:

BAB 1 PENDAHULUAN

Pada Bab 1 akan dijelaskan tentang latar belakang masalah yang diambil, maksud dan tujuan, batasan masalah, metodologi penelitian dan sistematika penulisan.

BAB 2 TINJAUAN PUSTAKA

Pada Bab 2 akan dijelaskan tentang tinjauan umum tempat penelitian dan pembahasan berbagai konsep dasar mengenai analisis berita hoax, *Neighbor Weighted K-Nearest Neighbor*, teori-teori pendukung lainnya yang berkaitan dengan topik pembangunan perangkat lunak.

BAB 3 ANALISIS MASALAH

Pada Bab 3 ini berisi tentang analisis dan perancangan yang dilakukan yang meliputi analisis masalah, analisis data masukan, analisis preprocessing, analisis pembobotan, analisis algoritma NWKNN, analisis non fungsional dan fungsional.

BAB 4 IMPLEMENTASI DAN PENGUJIAN SISTEM

Pada Bab 4 ini berisi tentang gambaran dari hasil implementasi dan pengujian pada sistem yang mulai pada proses pengklasifikasian oleh metode NWKNN dan melakukan tahapan pengujian dengan metode pengujian yang akan menampilkan hasil keluaran seperti nilai *Accuracy*, *Recall*, *Precision* dan *F-1 Score* pada sistem aplikasi yang telah diprogram untuk menguji akurasi dari analisis berita hoax pada berita hoax di Indonesia di situ Kaggle.

BAB 5 KESIMPULAN DAN SARAN

Pada Bab 5 akan dijelaskan tentang kesimpulan yang sudah diperoleh dari hasil penulisan dan saran agar dapat dilakukan pengembangan untuk penelitian selanjutnya.