

BAB I PENDAHULUAN

1.1 Latar Belakang

Media massa mengalami perkembangan dalam proses penyampaian pesan. Untuk saat ini proses penyampiannya dapat dilakukan seketika dengan melalui televisi, radio, atau dengan media on-line. Perkembangan proses komunikasi ini menggeser bentuk penyampaian berita pada wadah yang baru secara on-line, yaitu melalui media sosial. Penggunaan media sosial sebagai media baru untuk penyebaran berita mulai digemari. Hal ini merupakan dampak dari perkembangan teknologi yang mempengaruhi gaya hidup masyarakat dalam bersosialisasi [1]. Salah satu media sosial yang berkembang pesat adalah *Twitter*.

Twitter merupakan platform digital yang digunakan para penggunanya untuk menulis, menyebarkan berita disertai link dan juga gambar. *Twitter* dibentuk oleh Jack Dorsey bersama Noah Glass, Biz Stone, dan Evan Williams pada tahun 2006. Penelitian yang dilakukan oleh PT. Bakrie Telecom, terdapat 19,5 juta pengguna *Twitter*. Pada awal-awal tahun terbentuknya *Twitter*, pengguna dapat mengirim dan membaca 140 karakter dalam setiap kiriman. Hingga saat ini telah mengalami perkembangan yang mampu mengirim dan membaca 280 karakter dalam setiap postingan. Media sosial *Twitter* memiliki beragam fitur seperti *like*, *retweet*, *quote tweet*, dan *share*. Melalui media *Twitter*, pengguna dapat melakukan interaksi lain seperti memantau isu yang sedang ramai diperbincangkan (*trending topic*) dalam tingkat nasional bahkan dunia, serta dapat menulis (*tweet*) dengan menggunakan tagar (*hashtag*). Dengan adanya kontrol tersebut memberi kemudahan bagi oknum penyebar *hoax* untuk melancarkan aksinya.

Kata *hoax* berasal dari bahasa Inggris yang berarti kebohongan. Berdasarkan Kamus Besar Bahasa Indonesia (KBBI), *hoax* (*hoaks*) berarti informasi bohong. Berita *hoax* yang beredar luas di media sosial memiliki berbagai macam bentuk dan bahasa yang sulit untuk dibedakan dengan berita yang lebih kredibel [2]. Berita *hoax* yang tersebar dapat memperparah konflik-konflik yang berada pada media sosial, minimnya pengawasan dan kebebasan berekspresi dalam media sosial mengubah cara penyampaian dan penerimaan informasi oleh

masyarakat [3]. Menurut [4], sebuah berita atau informasi dengan isi yang provokatif atau bermuatan pendapat yang mudah membangkitkan emosi bagi pembacanya, mengundang ketakutan bagi pembacanya terhadap suatu hal yang terjadi, merupakan ciri-ciri umum sebuah berita atau informasi yang mengandung unsur *hoax*. Oleh karena itu dibutuhkan pengklasifikasian dengan tujuan menyaring berita agar masyarakat dapat membedakan berita *hoax* dan yang bukan.

Klasifikasi teks dapat digunakan sebagai salah satu sarana pendukung dalam pembangunan teknologi tersebut. Klasifikasi merupakan metode yang digunakan untuk pencarian dan penyaringan sebuah kelas data dengan tujuan dapat digunakan untuk melakukan prediksi kelas dari objek yang belum diketahui kelasnya [5]. Klasifikasi dokumen dapat diartikan proses menemukan kumpulan model yang mendeskripsikan serta membedakan kelas data sesuai kategori yang dimilikinya [6]. Pada penelitian ini, klasifikasi digunakan untuk mengklasifikasikan data yang telah didapatkan yaitu berupa cuitan (*tweet*) menjadi dua kelas, yaitu kelas berita *hoax* dan bukan *hoax*. Untuk membantu mengukur tingkat akurasi keberhasilan program tersebut, dalam penelitian ini digunakan metode *Naïve Bayes Classifier*.

Metode *Naïve Bayes Classifier* dikemukakan oleh Thomas Bayes dengan memanfaatkan perhitungan probabilitas dan statistik dalam melakukan klasifikasi sebuah data. *Naïve Bayes Classifier* membantu pengklasifikasian teks dengan menggunakan atribut kata yang muncul dalam satu dokumen untuk dasar pengklasifikasiannya [7]. Pemilihan metode ini untuk melakukan klasifikasi adalah karena memiliki beberapa kelebihan antara lain, memiliki tingkat akurasi tinggi. Pada penelitian [8], tingkat akurasi yang dihasilkan oleh *Naïve Bayes Classifier* lebih unggul apabila dibandingkan dengan algoritma lain (*K-Nearest Neighbor*) pada proses klasifikasi artikel, dengan tingkat presentase akurasi sebesar 70% berbanding 30% terhadap 40 data uji coba. Unggulnya *Naïve Bayes Classifier* karena salah satu kelebihan metode tersebut dalam melakukan klasifikasi terhadap data *training* dengan perhitungan parameter yang sedikit. Berdasarkan latar belakang yang telah diuraikan diatas, maka penulis melakukan penelitian

dengan judul “Klasifikasi Berita Hoax pada Media Sosial Twitter dengan Metode Naïve Bayes Classifier”.

1.2 Rumusan Masalah

Berdasarkan permasalahan pada latar belakang penulis mengambil rumusan masalah, yaitu bagaimana mengklasifikasi berita *hoax* yang pada media sosial *Twitter* dengan metode *Naïve Bayes Classifier*.

1.3 Batasan Masalah

Batasan masalah yang dibutuhkan sebagai batas permasalahan serta pembahasan pada penelitian ini dan tercapainya tujuan yang diharapkan. Berikut adalah batasan masalah pada penelitian:

1. Dataset yang digunakan berjumlah 500 tweet dengan kelompok berita umum.
2. Data yang diambil hanya teks yang menggunakan Bahasa Indonesia.
3. Klasifikasi yang dilakukan akan menampilkan 2 hasil, yaitu berita *hoax* dan bukan *hoax*.

1.4 Maksud dan Tujuan Penelitian

1. Membangun dan mengimplementasikan metode *Naïve Bayes Classifier* untuk membangun sistem klasifikasi berita *hoax* berbasis *web*.
2. Membantu memecahkan masalah klasifikasi berita *hoax* yang tersebar pada media sosial *Twitter*.

1.5 Manfaat Penelitian

Adapun manfaat penelitian ini adalah sebagai berikut:

1. Menambah pengetahuan tentang pengimplementasian metode *Naïve Bayes Classifier* pada sistem klasifikasi berita *hoax* berbasis *web*.

2. Dapat membantu peneliti selanjutnya untuk melakukan klasifikasi berita *hoax* dengan metode *Naïve Bayes Classifier* dengan tingkat akurasi yang maksimal.

1.6 Metode Penelitian

Untuk mencapai keberhasilan pada penelitian ini, terdapat beberapa metode yang digunakan oleh penulis sebagai acuan, yaitu adalah:

1.6.1 Metode Pengumpulan Data

Pada penelitian ini metode pengumpulan dataset dilakukan dengan cara manual. Pencarian data dimulai dari rentang waktu Desember 2019 sampai Desember 2020. Dataset yang dibutuhkan berupa cuitan yang disebar oleh pengguna media sosial *Twitter*. Sedangkan dataset didapatkan dari *Twitter API (Application Programming Interface)*. Tweet yang sudah dipilih akan dimasukkan pada database.

1.6.2 Metode Analisis

Data atau informasi yang telah didapatkan kemudian dipelajari dan dianalisa. Untuk mengetahui model yang direncanakan maka perlu dilakukan implementasi dari model ke bentuk program komputer. Analisis disini dilakukan pada data yang diperoleh dari kicauan (*tweet*) kemudian diolah dengan metode *Naïve Bayes Classifier*.

1.6.3 Metode Perancangan

Metode perancangan berisi penjelasan langkah-langkah pada proses perancangan aplikasi setelah dilakukan proses analisa. Langkah-langkah ini meliputi latar belakang atau ide perancangan, identifikasi masalah, tujuan perancangan, dan konsep perancangan dengan menggunakan UML. Perancangan sistem klasifikasi berita *hoax* dengan metode *Naïve Bayes Classifier* diawali dengan merubah kata yang terkandung dalam tweet kedalam kata dasar dengan menghilangkan kata imbuhan, penghubung,

awalah, dan akhirnya kemudian mencocokkan kata dan menghitung presentase kemiripan.

1.6.4 Metode Implementasi

Metode Implementasi adalah pengumpulan data yang diuji dan data latihan kemudian diolah sehingga menghasilkan pengembangan sistem dengan materi yang telah dirancang.

1.6.5 Metode Testing

Metode testing atau pengujian merupakan proses yang dilakukan dengan tujuan untuk memastikan fungsi-fungsi dari sistem bekerja dengan baik serta mencari kesalahan yang mungkin terjadi pada sistem. Penelitian ini diuji dengan metode *black box* dan *white box testing*. Proses pengujian dilakukan terhadap seluruh kebutuhan fungsional yang telah dirancang pada tahap perancangan sistem, untuk menemukan kesalahan yang dapat terjadi pada sistem yang dibangun, agar program yang dibangun dapat memenuhi kebutuhan fungsionalnya. Pengujian algoritma *Naïve Bayes* untuk mengukur tingkat akurasi klasifikasi dilakukan dengan matriks konfusi (*confusion matrix*).

1.7 Sistematika Penulisan

Penelitian ini terdiri dari enam bab, dengan susunan sebagai berikut:

BAB I PENDAHULUAN

Bab pendahuluan sebagian besar berisi latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan laporan penelitian.

BAB II LANDASAN TEORI

Bab Landasan Teori berisi tinjauan pustaka, menguraikan teori-teori pendukung judul penelitian dan mendasari pembahasan secara detail.

Landasan teori dapat berupa definisi atau model yang berkaitan dengan permasalahan yang diteliti.

BAB III ANALISA DAN PERANCANGAN

Bab Analisa dan Perancangan berisi analisa kebutuhan data dan pelabelan data, analisa proses *text pre-processing*, pembobotan fitur dan analisa metode *Naïve Bayes Classifier* serta perancangan aplikasi.

BAB IV IMPLEMENTASI DAN PENGUJIAN

Bab ini berisi tentang hasil implementasi dari perancangan yang telah dilakukan serta pengujian terhadap aplikasi.

BAB V PENUTUP

Bab ini berisi tentang kesimpulan dan saran. Kesimpulan dapat mengemukakan kembali masalah penelitian, menyimpulkan bukti-bukti yang diperoleh dan menarik kesimpulan apakah hasil yang didapat mampu untuk diimplementasikan.

