

BAB I

PENDAHULUAN

1.1 Latar Belakang

Adanya Era Revolusi 4.0 penggunaan teknologi informasi semakin meningkat drastis. Banyak penduduk di dunia dengan mudah berkomunikasi satu dengan lainnya hanya dengan hitungan detik, menyebabkan komunikasi adalah kebutuhan utama dalam kehidupan manusia. Penyampaian pesan ini dengan mudah masyarakat terima dari berbagai media sosial seperti *Facebook*, *Twitter*, *Instagram* dan pesan elektronik lainnya seperti *Whatsapp* dan *Telegram* tanpa mengetahui kualitas pada berita baik dan buruk yang sedang dikonsumsi [1].

Berdasarkan data dari *Global Web Index* di tahun 2016 dari 7,39 miliar manusia, sekitar 31% diantaranya atau 2,31 miliar manusia di muka bumi ini adalah pengguna aktif media sosial. Dengan lama penggunaan untuk wilayah Indonesia menduduki peringkat 9 atau sekitar 2,9 jam/perhari. Sedangkan penduduk dunia yang menghabiskan waktu paling lama di media sosial adalah Filipina dengan data – rata lama penggunaan sebanyak 3,7 jam/perhari. Untuk jenis layanan kelompok saling bertikai dengan dasara ‘Pra sangka buta’ [2]. Banyaknya penggunaan media sosial di dunia tersebut mempunyai banyak tujuan positif diantaranya untuk menjalin persahabatan, perdagangan, penyampaian kejadian terkini, dan berbagai informasi lainnya. Hal ini tidak menutup kemungkinan penduduk dunia dengan sengaja melakukan berbagai aksi-aksi tindak kejahatan yang dapat menyebabkan kerugian sekelompok orang, suku, organisasi lainya atau kita sebut *cybercrime*.

Di Indonesia, selain beberapa kasus di atas negara ini juga sedang dihadapkan dengan 'krisis berita' berupa penyebaran berita *hoax* yang merambah secara masif dan terstruktur dari level sosial atas hingga level sosial bawah, dari kota hingga ke desa, dan bahkan dari kalangan tua hingga ke kalangan muda [2]. Akibat terjadinya berita *hoax* yang sedang dikonsumsi oleh penduduk dunia tanpa mengetahui sumber awal dapat menyebabkan persulutan amarah hingga permusuhan berbagai golongan.

Facebook adalah jejaring sosial media yang terkenal saat ini untuk memberikan informasi terkini yang dapat dikonsumsi oleh publik. Setiap pengguna *Facebook* bebas mengunggah posting-an atau komentar tanpa ada batasan. Dengan keberadaan interaksi ini akan memberikan efek interaksi sosial dan membentuk semacam jejaring pertemanan [3]. Oleh sebab itu dibutuhkan sebuah analisis yang tepat guna mengelola data teks sehingga menyuguhkan informasi pesan yang berharga mengenai informasi yang disampaikan dari berbagai pengguna *facebook*. Agar pesan tersebut dapat dimanfaatkan dan berguna, dibutuhkan berbagai proses sehingga didapatkan suatu informasi yang penting melalui analisis sentimen [4].

Analisis sentimen disebut juga dengan *opinion mining* (penambangan opini) yaitu proses untuk mengekstrak suatu opini atau pendapat dari dokumen untuk topik tertentu [4]. Analisis sentimen juga biasa digunakan untuk mengetahui tingkat emosional suatu masalah atau peristiwa terhadap teks apakah cenderung hal negatif atau positif. Teknik yang digunakan ini disebut *Text Mining*. *Text mining* merupakan sebuah teknik yang digunakan untuk mengekstraksi informasi yang berguna dari data teks yang tidak terstruktur.

Text mining mengekstraksi kata kunci atau mengekstraksi pendapat dan ulasan analisis text sehingga dapat mendukung untuk memahami pendapat masyarakat dalam data text [4]. Sebelum dilakukan klasifikasi, perlu tahap *preprocessing*. *Preprocessing* mempunyai tahap sebagai berikut yaitu *case folding*, *tokenzing*, *stopwhatch removal* dan *steaming*. Proses *steaming* menjadi tahapan paling penting di dalam tahap *preprocessing* dikarenakan pada *steaming* terjadi proses penghilangan kata imbuhan, sehingga menjadi kata dasar. Dalam penelitian data sosial media lain seperti *Twitter* digunakan untuk analisis sentimen opini Pilkada DKI 2017 dengan menggunakan metode *Naïve Bayes Classifier* menghasilkan nilai akurasi tertinggi sebesar 74.81%. Penelitian berjudul "*Indonesian News Classification based on NaBaNA*" penelitian ini mengategorikan berita berbahasa Indonesia pada aplikasi berita Jawa Tengah menggunakan *Naïve Bayes Classifier* dan *stemming* dari Naizief-Adriani. Penelitian ini menghasilkan nilai akurasi sebanyak 94% [4]. Metode yang digunakan untuk mengidentifikasi *tweet* hoaks yang berhubungan dengan pemilihan presiden 2019 adalah *Naïve Bayes Classifier (NBC)*. Metode *NBC* dipilih karena dapat mengidentifikasi berita hoaks dengan akurasi sebesar 91.36% [5].

Dari referensi yang ada, maka dapat disimpulkan bahwa *Naïve Bayes Classifier* merupakan salah satu metode yang memiliki akurasi tinggi yang dapat digunakan untuk peneliti dalam melakukan analisis. Selain itu, *NBC* telah terbukti efektif untuk kategorisasi teks, prosesnya sederhana, cepat dan akurasi klasifikasi yang tinggi [5]. Keuntungan penggunaan *Naive Bayes* adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (*Training Data*) yang

kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Naive Bayes sering bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks dari pada yang diharapkan [6]. Berdasarkan uraian di atas, maka penulis akan melakukan penelitian sebagai skripsi dengan judul **“Analisis Pesan Hoax Pada Facebook Menggunakan Naive Bayes Classifier”**.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan diatas, maka penulis dapat merumuskan rumusan masalah sebagai berikut :

1. Bagaimana penerapan algoritma *Naive Bayes Classifier* dalam mendeteksi pesan *hoax* di media sosial *Facebook*?
2. Bagaimana kualitas tingkat akurasi dari algoritma *Naive Bayes Classifier*?

1.3 Batasan Masalah

Supaya penelitian ini dapat terfokus maka diperlukan adanya batasan-batasan untuk membatasi ruang lingkup dari penelitian ini. Adapun batasan-batasan dari penelitian ini adalah sebagai berikut :

1. Penggunaan metode algoritma klasifikasi yaitu *Naive Bayes Classifier* untuk mengkasifikasi pesan informasi yang bersifat *hoax* dan fakta.
2. Pesan maupun kalimat yang digunakan menggunakan bahasa utama yaitu bahasa Indonesia.
3. Pengolahan dan pelabelan data dilakukan secara manual berdasarkan studi literatur dari berbagai sumber.

4. Sistem yang digunakan dan dibangun hanya berbasis desktop.
5. Pengembangan aplikasi dan model klasifikasi dibangun menggunakan Bahasa pemrograman *Python*.

1.4 Maksud dan Tujuan Penelitian

Maksud dan tujuan dari penelitian ini sebagai berikut :

1. Membangun sistem yang dapat mengklasifikasi pesan yang merujuk ke arah *hoax* menggunakan metode *Naïve Bayes Classifier*.
2. Mengetahui tingkat akurasi dari hasil klasifikasi pada algoritme *Naïve Bayes Classifier*.

1.5 Manfaat Penelitian

Adapun manfaat yang diharapkan dari penelitian ini diantaranya sebagai berikut :

- a. Menambah wawasan dan pemahaman secara mendalam terkait metode *Naïve Bayes Classifier* untuk menganalisis sebuah pesan apakah sesuai dengan fakta atau tidak (*hoax*).
- b. Bagaimana mendapatkan hasil klasifikasi yang dapat dipercaya, artinya nilai akurasinya tinggi (>90%).

1.6 Metode Penelitian

Untuk mendukung penelitian ini, penulis menggunakan beberapa metodologi penelitian diantaranya adalah :

Metode penumpulan data merupakan metode yang diperlukan untuk memperoleh informasi yang dibutuhkan untuk menunjang penelitian. Adapun

metode pengumpulan data yang akan di gunakan dalam penelitian ini adalah metode studi pustaka atau *crawling data*.

1.6.1 Studi Pustaka

Studi pustaka dilakukan sebagai proses menghimpun informasi yang relevan terhadap topik atau permasalahan dari buku-buku, karya ilmiah, jurnal-jurnal, skripsi maupun situs internet yang dapat membantu penelitian. Hal ini penulis mencari informasi mengenai *Naïve Bayes Classifier*, *K-Folf Cross Validation*, *Facebook*, *Teori Hoax media sosial*, *Berita hoax*, dan informasi yang relefan lainnya.

1.6.2 Metode Crawling Data

Pengambilan data yang dilakukan akan menggunakan *crawling* memakai *tool Fanspage* dengan memanfaatkan akses token dari *Facebook*. Metode ini memungkinkan untuk memanggil data berupa status atau komentar *facebook* yang di inginkan pada halaman *fanspage Facebook* terhadap pesan berita yang akan dianalisis sebagai sumber data. Kriteria yang dibutuhkan menghindari komentar yang berisi emoticon dan karakter lainnya. Jumlah data yang diperlukan harus memenuhi target yang telah ditentukan. Selanjutnya dilakukan proses data sampai menghasilkan *dataset* yang siap diklasifikasi.

1.6.3 Metode Testing

Metode testing yang digunakan dalam penelitian ini menggunakan *Black Box Testing* dan pengujian model menggunakan *K-Fold Cross Validation* dan *Confusion Matrix*.

1.7 Sistematika Penulisan

Sistematika penulis yang akan digunakan oleh penulis dalam penelitian ini yaitu membagi setiap bab menjadi 5 bab dengan tujuan supaya memudahkan penulis dalam melakukan penyusunan skripsi dan memberikan kenyamanan bagi pembacanya. Sistematika penulisan yang dibuat adalah sebagai berikut :

BAB I PENDAHULUAN

Pada bab ini berisi mengenai latar belakang penelitian, rumusan penelitian, batasan penelitian, maksud dan tujuan penelitian, metode penelitian yang digunakan serta sistematika penulisan penelitian.

BAB II LANDASAN TEORI

Bab Landasan teori akan menyajikan berbagai teori yang diperoleh oleh berbagai sumber seperti buku-buku, karya ilmiah, jurnal-jurnal, skripsi maupun situs internet sebagai bentuk referensi dalam membantu pemahaman terkait penelitian.

BAB III ANALISIS DAN PERANCANGAN

Bab analisis dan perancangan menyajikan berbagai kebutuhan sistem yang terkait dengan objek, mulai dari teknik pengumpulan data, menganalisa kebutuhan dan kesalahan dalam proses penyusunan sistem.

BAB IV IMPLEMENTASI DAN PEMBAHASAN

Bab implementasi dan pembahasan ini akan menyajikan pembahasan terkait hasil dari penelitian yang sudah melakukan berbagai tahapan

BAB V PENUTUP

Bab penutup akan menyajikan hasil kesimpulan yang diperoleh dari tahap-tahap yang dilakukan selama penelitian yang disertai dengan saran pengembangan demi kemajuan sistem di waktu yang akan datang.

