

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Di masa sekarang ini, teknologi telah berkembang dengan sangat pesat. Teknologi hadir dengan tujuan untuk membantu memudahkan manusia dalam melakukan pekerjaannya sehari-hari. Salah satu bidang yang merasakan perkembangan dari teknologi adalah pada bidang jurnalistik yang dimana sebelumnya media berita hanya menggunakan media cetak dalam menyebarkan berita dan orang-orang jika ingin membaca berita harus melalui koran atau media cetak lainnya. Namun dengan adanya perkembangan dari teknologi, maka sekarang media berita dapat menyebarkan berita secara daring salah satunya melalui *website* dan orang-orang dapat membaca berita dimanapun dan kapanpun dengan akses yang mudah dan gratis. Berita sendiri merupakan bentuk laporan tentang suatu kejadian yang sedang terjadi baru-baru ini atau keterangan terbaru dari suatu peristiwa. Dengan kata lain berita adalah fakta menarik atau sesuatu hal yang penting yang disampaikan pada masyarakat orang banyak melalui media [1].

Dalam jurnalistik, bahasa digunakan untuk menyampaikan informasi yang tepat dan akurat kepada masyarakat melalui media massa [2]. Salah satu fungsi dari bahasa adalah sebagai alat untuk melakukan komunikasi sehingga dalam penggunaan bahasa khususnya bahasa Indonesia haruslah menggunakan ejaan yang baik dan benar berdasarkan kaidah yang ada. Ejaan sendiri merupakan tata cara penggunaan kata, kalimat, dan tanda baca baik dalam bentuk lisan maupun tertulis [3].

Perkembangan teknologi di bidang jurnalistik telah berkembang dengan sangat pesat, namun masih sering terjadi penyimpangan terhadap bahasa pada portal berita daring. Biasanya, hal ini terlihat dari aspek pemakaian ejaan dan kata yang tidak sesuai dengan kaidah penulisan yang sudah ditetapkan dan hal tersebut tidak jarang terjadi pada portal berita daring [4]. Hal ini dapat terjadi dikarenakan kecepatan dari penyebaran berita itu sendiri sehingga biasanya menyebabkan kesalahan pada saat melakukan pengetikan berita dan juga ketika sedang dalam proses penyuntingan. Kesalahan ejaan yang terjadi pada berita dapat menyebabkan informasi yang terdapat dalam berita menjadi tidak jelas dan ambigu [5]. Pada penelitian yang pernah dilakukan terkait analisis kesalahan berbahasa pada portal

berita daring dengan studi kasus portal berita daring suara.com dan penelitian terkait kesalahan penulisan kata dan tanda baca pada berita daring, didapat kesimpulan bahwa masih banyak ditemukan kesalahan dalam pengetikan pada berita [6][7].

Berdasarkan permasalahan yang ada, dilakukanlah penelitian untuk mendeteksi kesalahan tik pada portal berita daring yang dimana pada penelitian ini akan menggunakan berita dari portal berita Tribunnews. Hal ini dikarenakan terdapat permintaan dari mitra terkait pembuatan beberapa modul dan salah satunya adalah modul untuk mendeteksi kesalahan tik. Terdapat alasan mengapa Tribunnews ingin membuat modul ini, yaitu dikarenakan Tribunnews mengunggah 3000 sampai 5000 artikel dalam sehari dan setiap reporter dituntut untuk menulis 20 artikel. Dengan jumlah produksi yang sebegitu besar, kemungkinan untuk terjadi kesalahan dalam penerbitan menjadi besar. Kesalahan-kesalahan dalam berbahasa pada berita dapat membuat kredibilitas dan kepercayaan masyarakat terhadap informasi yang disajikan media menurun [8]. Tribunnews sendiri merupakan portal berita daring nomor satu di Indonesia yang dikelola oleh PT Tribun Digital Online. Tribunnews.com memiliki jaringan yang telah tersebar ke seluruh penjuru Indonesia yang bernama Tribun Network [9].

Adapun beberapa penelitian yang pernah dilakukan oleh mahasiswa dari Universitas Multimedia Nusantara mengenai deteksi kesalahan kata yang diharapkan dapat membantu penelitian ini dalam memperoleh wawasan tambahan dalam melakukan penelitian deteksi kesalahan tik. Penelitian yang pernah dilakukan adalah deteksi kesalahan eja kata luluh menggunakan algoritma *jaccard similarity* yang dimana penelitian ini didapati tingkat akurasi sebesar 93.2% [10]. Kemudian terdapat juga penelitian mengenai deteksi penggunaan kata konjungsi menggunakan algoritma *consine similarity* yang dimana pada penelitian ini didapati tingkat akurasi sebesar 92.2% [11].

Penelitian pendeteksian kesalahan tik ini akan menggunakan algoritma *Random Forest*. Alasan mengapa penelitian ini menggunakan algoritma *Random Forest* dikarenakan *random forest* merupakan salah satu algoritma dalam *ensemble learning* yang digunakan untuk melakukan pengklasifikasian data set yang berjumlah besar. Selain itu terdapat beberapa kelebihan dari algoritma ini, seperti memiliki hasil akurasi yang bagus, relatif kuat terhadap *outliers* dan *noise*, sifatnya yang sederhana dan mudah dipararelkan [12]. Dalam beberapa penelitian yang pernah dilakukan terkait pengklasifikasian teks juga menunjukkan bahwa algoritma *Random Forest* memiliki tingkat akurasi yang cukup tinggi seperti analisis sentimen terhadap aplikasi Ruangguru yang memiliki tingkat akurasi sebesar 97,16% [12]

dan deteksi berita palsu yang memiliki tingkat akurasi sebesar 84% [13].

Model deteksi yang dibuat dapat mendeteksi dengan melakukan pembelajaran dari data baru yang telah dilatih sehingga model dapat menambah pengetahuannya sehingga tidak mendeteksi dengan mencocokkan kata atau *string matching*. Selain itu, model yang dibuat tidak hanya sampai pada deteksi kesalahan tik saja, namun model dapat memberikan saran koreksi kata yang benar dari kata salah tik yang telah terdeteksi sebelumnya oleh model.

Dengan dilakukannya penelitian ini, diharapkan dapat melakukan pendeteksian kesalahan tik dengan baik dan benar sehingga dapat membantu penulis berita dalam memeriksa kesalahan tik pada artikel berita yang ditulis. Kemudian juga dengan berhasilnya penelitian ini diharapkan informasi yang terdapat pada artikel berita yang ditulis dapat tersampaikan dengan baik kepada pembaca.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang yang ada, maka masalah yang dapat dirumuskan pada penelitian ini adalah sebagai berikut.

1. Bagaimana cara mendeteksi kesalahan tik menggunakan algoritma *Random Forest* pada artikel berita Tribunnews?
2. Bagaimana tingkat akurasi dari model deteksi kesalahan tik pada artikel berita Tribunnews?

## **1.3 Batasan Permasalahan**

Batasan-batasan masalah yang ada pada penelitian ini adalah sebagai berikut.

1. Dataset yang digunakan adalah artikel berita berbahasa Indonesia dari portal berita daring bernama Tribunnews dengan total sebanyak 50 artikel berita. 50 artikel berita digunakan sebagai sampel untuk menguji kemampuan model dalam melakukan deteksi kesalahan tik. Meskipun ukuran sampel yang digunakan belum mencakup seluruh artikel berita yang terdapat pada Tribunnews, namun penggunaan 50 artikel berita sebagai sampel ini dapat memberikan wawasan awal dan informasi yang berguna dalam mengetahui kemampuan model.

2. Kesalahan yang dideteksi hanya kata yang terdapat pada Pendoman Umum Ejaan Bahasa Indonesia (PUEBI) dan tidak termasuk angka, nama tempat, nama orang, nama organisasi, singkatan, dan kata selain bahasa Indonesia.

#### **1.4 Tujuan Penelitian**

Berdasarkan rumusan masalah yang ada, maka tujuan dari penelitian ini adalah sebagai berikut.

1. Mendeteksi kesalahan tik menggunakan algoritma *Random Forest* pada artikel berita berbahasa Indonesia.
2. Mengukur tingkat akurasi dari algoritma *Random Forest* yang digunakan pada model deteksi kesalahan tik pada artikel berita berbahasa Indonesia.

#### **1.5 Manfaat Penelitian**

Adapun manfaat yang dapat diambil dari penelitian ini, antara lain sebagai berikut.

1. Membantu dalam mendeteksi kesalahan tik pada artikel berita Tribunnews.
2. Mengetahui tingkat akurasi dari algoritma *Random Forest* yang digunakan pada model deteksi kesalahan tik pada artikel berita Tribunnews.

#### **1.6 Sistematika Penulisan**

Berisikan uraian singkat mengenai struktur isi penulisan laporan penelitian, dimulai dari Pendahuluan hingga Simpulan dan Saran.

Sistematika penulisan laporan adalah sebagai berikut:

- Bab 1 PENDAHULUAN

Bab ini berisikan latar belakang masalah, rumusan masalah, batasan permasalahan, tujuan penelitian, manfaat penelitian dan sistematika penulisan dari laporan penelitian ini.

- Bab 2 LANDASAN TEORI

Bab ini berisikan penjelasan dari teori-teori yang digunakan dalam penelitian ini.

- Bab 3 METODOLOGI PENELITIAN

Bab ini berisikan penjelasan mengenai metode-metode yang digunakan selama melakukan penelitian ini.

- Bab 4 HASIL DAN DISKUSI

Bab ini berisikan hasil implementasi dari algoritma *random forest*, hasil uji coba dan evaluasi dari model deteksi kesalahan tik yang dibuat pada penelitian ini.

- Bab 5 SIMPULAN DAN SARAN

Bab ini berisikan kesimpulan yang dapat diambil dari penelitian ini dan saran yang dapat digunakan untuk mengembangkan penelitian ini kedepannya.

