



### **Hak cipta dan penggunaan kembali:**

Lisensi ini mengizinkan setiap orang untuk menggubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

### **Copyright and reuse:**

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Kemajuan teknologi yang terus berkembang memberikan dampak yang cukup besar terhadap perkembangan bisnis informasi yang dulunya menggunakan media cetak dan kemudian berinovasi dengan beralih ke media elektronik. Hal ini juga yang menyebabkan munculnya berbagai macam pesaing baru dalam bisnis informasi yang dapat kita lihat dengan banyaknya media berita daring yang bermunculan. Salah satu media berita daring yang cukup populer adalah Tribunnews.com.[3]

Berdasarkan survey yang dilakukan oleh Reuters Institute masyarakat Indonesia mengkonsumsi media berita daring sebagai sumber berita utama yakni sebesar 88% dan Tribunnews.com menempati peringkat ke-4 sebagai media berita daring yang sering dikunjungi oleh masyarakat Indonesia dengan presentase 33,5% [4]. Dari data survey tersebut dapat diamati juga bahwa media berita daring merupakan sumber informasi yang cukup berpengaruh bagi masyarakat Indonesia. Namun, terdapat beberapa artikel yang terdapat pada berita daring yang ejaannya tidak sesuai dengan Pedoman Umum Ejaan Bahasa Indonesia(PUEBI). Kesalahan ejaan atau kesalahan ejaan pada artikel Tribunnews biasanya terjadi pada tanda baca dan kesalahan pemakaian tulisan[5]. Kesalahan ejaan ini dapat menyebabkan terjadinya kerancuan pada kalimat yang menyebabkan makna asli dari hal yang ingin disampaikan menjadi kurang jelas[6].

Kesalahan pemakaian tulisan tersebut pada umumnya disebabkan kurangnya ketelitian penulis atau kurangnya pengetahuan tentang PUEBI[7]. Kesalahan ejaan yang terdapat pada media berita elektronik dapat mengakibatkan pembaca yang membaca artikel tersebut tidak dapat memahami isi dari artikel tersebut ataupun mengakibatkan pembaca yang membaca artikel tersebut mendapatkan misinformasi baik itu dari segi isi artikel maupun dari segi ejaan. Berdasarkan penelitian lain dalam kasus serupa yang dibangun menggunakan metode *Cosine Similarity* untuk deteksi kesalahan konjungsi dapat menghasilkan ketepatan akurasi 92.2% dan penelitian tentang deteksi kesalahan kata lulu dengan algoritma *Jaccard Similarity* yang mencapai tingkat akurasi 93.2% [8][9].

Berdasarkan informasi yang telah didapat, penelitian ini bertujuan

untuk mendeteksi kesalahan tik yang terdapat pada artikel berita Tribunnews menggunakan metode *supervised learning* dengan menggunakan algoritma *Multinomial Naive Bayes*. Model yang dibuat menggunakan algoritma ini nantinya dapat digunakan untuk mendeteksi kesalahan tik secara otomatis pada artikel berita Tribunnews. *Multinomial Naive Bayes* dipilih sebagai algoritma yang digunakan karena berdasarkan penelitian sebelumnya yang juga dilakukan untuk mengklasifikasikan artikel berita *Multinomial Naive Bayes* memiliki *f-measures* sebesar sebesar 95% [10] dan *Multinomial Naive Bayes* menunjukkan kemampuan pemrosesan data yang cepat dibandingkan algoritma NLP lain yakni mencapai 5.4 kali lebih cepat [11]. Hal ini dapat juga dilihat dari hasil perbandingan *Multinomial Naive Bayes* dengan algoritma *K-Nearest Neighbors* dan *Convolutional Neural Network* dalam mendeteksi kesalahan tik bahasa Inggris dimana *Multinomial Naive Bayes* memiliki hasil yang lebih unggul dalam hal akurasi dan kecepatan komputasi dibandingkan algoritma *K-Nearest Neighbors* dan *Convolutional Neural Network* [12]. Terdapat juga penelitian sebelumnya yang membandingkan *Multinomial Naive Bayes* dan algoritma *Long Short Term Memory Network* untuk klasifikasi teks pada 2 bahasa yakni bahasa Inggris dan bahasa India, hasil dari penelitian ini menunjukkan bahwa *Multinomial Naive Bayes* lebih unggul dalam hal klasifikasi teks bahasa Inggris [13], oleh karena itu algoritma *Multinomial Naive Bayes* ini dipilih pada penelitian ini. Namun, dalam penelitian ini, pada tahap *preprocessing* data untuk bahasa Indonesia tidak dilakukan proses *stemming* atau penghapusan imbuhan.

Hal ini dikarenakan penggunaan *stemming* dapat menghilangkan imbuhan yang dapat mempengaruhi makna dan kelengkapan kata pada artikel. Oleh karena itu, dalam penelitian ini, huruf berimbuhan akan tetap dipertahankan untuk menjaga kelengkapan kata pada artikel dan mempertahankan informasi yang relevan dalam proses klasifikasi dengan menggunakan algoritma *Multinomial Naive Bayes*. Penelitian ini diharapkan dapat membantu para penulis dalam hal mendeteksi kesalahan ketik dan juga mempersingkat waktu untuk pengecekan ulang artikel.

## 1.2 Rumusan Masalah

Dari latar belakang yang telah dijelaskan, maka rumusan masalah yang akan dijawab dalam penelitian ini adalah sebagai berikut:

1. Bagaimana menerapkan algoritma *Multinomial Naive Bayes* untuk mendeteksi kesalahan tik pada artikel berita Tribunnews?

2. Bagaimana cara mengukur tingkat akurasi dari model yang dibuat menggunakan *precision*, *recall* dan *f1-scores*.

### 1.3 Batasan Permasalahan

Adapun batasan masalah yang akan ditetapkan dalam penelitian ini adalah sebagai berikut:

1. Dataset yang digunakan dalam penelitian ini didapatkan dari artikel berita Tribunnews, dataset diperoleh dengan cara *scraping* dari beberapa situs dan juga didapatkan dari menginputnya secara manual dengan jumlah dataset sekurang-kurangnya 50000 baris dan 2 buah kolom yakni: kolom kata dan kolom label kata. Data artikel Tribunnews sebanyak 50 artikel yang di terbitkan dari tahun 2021 sampai dengan 2022.
2. Artikel yang digunakan menggunakan bahasa Indonesia.
3. Kesalahan yang dideteksi dibatasi oleh kata yang terdapat pada PUEBI dan tidak mendeteksi kesalahan penulisan nama tempat ataupun nama orang.

### 1.4 Tujuan Penelitian

Tujuan dari penelitian ini sebagai berikut:

1. Menerapkan algoritma *Multinomial Naive Bayes* untuk mendeteksi kesalahan tik pada artikel berita Tribunnews.
2. Menerapkan algoritma *Multinomial Naive Bayes* dengan memanfaatkan seluruh data latih yang ada.

### 1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah untuk mendeteksi kesalahan tik pada artikel berita Tribunnews secara otomatis sehingga dapat menghemat waktu para penulis artikel dalam mempublikasikan hasil tulisannya.

### 1.6 Sistematika Penulisan

- Bab 1 PENDAHULUAN, pada bagian ini menjelaskan permasalahan yang diteliti pada penelitian ini, bagian ini dibagi dalam beberapa bagian kecil

seperti latar belakang penelitian, rumusan masalah, batasan masalah, tujuan penelitian dan manfaat dari penelitian.

- Bab 2 LANDASAN TEORI, bagian ini memuat beberapa landasan teori pendukung yang dipakai pada penelitian antara lain, *Natural Language Processing*, kesalahan tik, *Text Preprocessing*, *Web Crawling* dan *Multinomial Naive Bayes*
- Bab 3 METODOLOGI PENELITIAN, pada bagian ini menjelaskan tentang langkah-langkah kerja penelitian serta rancangan sistem yang dibuat.
- Bab 4 HASIL DAN DISKUSI, pada bagian ini menjelaskan hasil implementasi dari langkah-langkah kerja penelitian yang dilakukan dimulai dari proses pengumpulan data sampai dengan evaluasi hasil.
- Bab 5 SIMPULAN DAN SARAN, pada bagian ini menjelaskan kesimpulan akhir dari penelitian yang dilakukan serta saran untuk pengembangan penelitian-penelitian lanjutannya.

