

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Bahasa adalah suatu komunikasi yang digunakan oleh manusia untuk melakukan interaksi dengan manusia lainnya. Bahasa ini banyak digunakan pada media sosial, televisi dan dimanapun pada saat berinteraksi dengan orang lain [1]. Selain itu terdapat juga hasil riset yang dilakukan oleh Kementerian Komunikasi dan Informatika bersama dengan Katadata yang menunjukkan sumber berita yang paling dipercaya adalah televisi dan media sosial menjadi dua sumber berita yang paling terpercaya oleh masyarakat Indonesia. Berdasarkan data yang didapat maka dapat diketahui bahwa media sosial menjadi peranan yang penting dalam penyebaran berita di Indonesia. Selain itu juga berita harus tidak melakukan kesalahan dalam segi pengejaan karena akan menyebabkan kesalahan pengertian sehingga masyarakat menanggapi berita dengan arti yang salah. Berita harusnya memiliki ciri khas yaitu menggunakan bahasa lugas dan juga berpedoman pada Pedoman Umum Ejaan Bahasa Indonesia [2].

Kata Keterangan adalah suatu kata yang mendampingi kata lainnya. Dalam hal ini kata keterangan berguna untuk memberi makna dan juga memberitahu agar dapat dimengerti kalimat tersebut sehingga memiliki pengaruh yang besar terhadap kalimat [3]. Dalam berita masih banyak ditemukan beberapa kesalahan dalam ejaan, namun terdapat juga kesalahan kata lainnya seperti segi tanda baca dan penulisan bilangan. Padahal berita harus ditulis secara sistematis, logis, dan komunikatif. Dari kesalahan ejaan ini maka bisa menyebabkan ketidaksesuaian fakta dan menjadi berita yang ambigu [4].

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

Menurut Rahman, Kata keterangan waktu adalah kata yang diperlukan untuk menyampaikan informasi yang sesuai dan berdasarkan fakta yang ada. Selain itu juga kata keterangan waktu harus mengikuti verba pewarta yang menunjukkan kalimat sebagai pemberitahuan informasi [5]. Menurut Alarisy, masih ditemui kesalahan kata pada berita yang dimana merupakan kesalahan yang fatal karena bisa dianggap menjadi bahasa yang baik atau tidak sesuai dengan kebenarannya [5].

Berdasarkan data - data yang telah dikumpulkan oleh similarweb, maka didapatkan kesimpulan dari hasil data yang terkumpul tentang Tribunnews.com. Tribunnews.com, memiliki tingkat rank 3 atau rank yang cukup tinggi dan selalu bertahan dalam beberapa tahun kebelakang terutama di bidang berita. Selain itu juga tribunnews.com lebih sering diakses untuk melakukan pembacaan berita sebanyak 72,75% berdasarkan traffic website tersebut [6]. Oleh karena itu maka didapatkan kesimpulan bahwa website tribunnews.com, lebih banyak diakses di bidang beritanya.

Terdapat juga beberapa penelitian terdahulu dalam hal *natural language processing* ini, beberapa penelitian terkait dengan deteksi bahasa indonesia dilakukan oleh Olwen [7] dan Evan [8] untuk mendeteksi kesalahan konjungsi dan peluluhan kata pada berita berbahasa Indonesia. Pada penelitian tersebut mendapat akurasi sebesar 92,2% untuk penelitian deteksi konjungsi dan akurasi sebesar 93,2% untuk deteksi keluluhan kata.

selain itu juga terdapat penelitian terdahulu terkait dengan teteksi kemiripan dokumen dengan penerapan *cosine similarity* dan pembobotan TF-IDF oleh Riyami. Pada penelitian tersebut mendapatkan akurasi sebesar 98% untuk deteksi kemiripan antara dua dokumen. Pada penelitian kemiripan antara dua dokumen, menunjukan kelebihan dari algoritma *cosine similarity* yang tidak terpengaruh dengan banyaknya data sehingga bisa mendeteksi data yang besar [9].

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

Berdasarkan penelitian terdahulu yang menunjukkan hasil - hasil yang memuaskan dalam pendeteksi kata atau kesamaan dalam dua dokumen, maka dipilihlah algoritma *cosine similartiy* dalam penelitian ini untuk mendeteksi kata keterangan waktu pada portal berita TribunNews. Terdapat juga salah satu kesalahan dalam protal berita Tribunnews yaitu tahun yang dimana seharusnya kata yang benar adalah tahun. Maka dari itu penelitian ini diharapkan memudahkan pengguna untuk melakukan deteksi kesalahan keterangan waktu pada portal berita TribunNews dan memperbaiki kesalahan keterangan waktu tersebut.

1.2 Rumusan Masalah

1. Bagaimana cara melakukan implementasi algoritma *Cosine Similarity* dalam mendeteksi keterangan waktu dalam portal berita TribunNews?
2. Bagaimana akurasi dari algoritma *cosine similarity* dalam menghitung kesalahan ejaan keterangan waktu pada portal berita TribunNews?
3. Bagaimana cara melakukan koreksi secara otomatis pada keterangan waktu dalam portal berita TribunNews?

1.3 Batasan Permasalahan

1. Proses Pendeteksi keterangan waktu pada portal berita hanya menggunakan bahasa Indonesia
2. Berita yang akan digunakan adalah berita dalam jangka 5 tahun dari portal berita Tribunnews
3. Pada Penelitian ini dapat mendeteksi maksimal dua gram keterangan waktu pada portal berita Tribunnews

1.4 Tujuan Penelitian

1. Melakukan implementasi algoritma *Cosine Similarity* dalam mendeteksi keterangan waktu dalam portal berita TribunNews
2. Mengetahui dan menghitung tingkat akurasi algoritma *cosine similarity* dalam penentuan keterangan waktu pada portal berita TribunNews dengan *confusion matrix*
3. Melakukan perbaikan secara otomatis dengan menampilkan hasil yang paling mirip di tabel *similarity* untuk menunjukkan perbaikan terhadap gram tersebut di portal berita Tribunnews

1.5 Manfaat Penelitian

1. Hasil penelitian ini dapat mendeteksi keterangan waktu yang terdapat pada portal berita TribunNews
2. Hasil penelitian ini dapat mengetahui akurasi algoritma *cosine similarity* pada keterangan waktu di portal berita TribunNews dengan *confusion matrix*
3. Hasil penelitian ini dapat memperbaiki 2-gram (bigram) yang salah pada keterangan waktu di portal berita TribunNews



1.6 Sistematika Penulisan

Sistematika penulisan ditulis dalam 5 bab yaitu BAB 1, BAB 2, BAB 3, BAB 4, dan BAB 5. Berikut ini penjabaran bab tersebut.

BAB 1 Pendahuluan Pada bab ini berisi tentang masalah yang diangkat pada laporan skripsi, yang dijelaskan melalui latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB 2 Landasan Teori Pada bab ini berisi mengenai landasan teori yang digunakan dalam penyusunan laporan skripsi yang berkaitan dengan topik yang dibahas dalam penelitian ini, yaitu *natural language processing*, *text preprocessing*, jarak, *streamlit*, *cosine similarity*, dan *confusion matrix*

BAB 3 Metodologi Penelitian Pada bab ini berisi tentang metodologi penelitian yang dilakukan dalam penelitian ini, serta perancangan yaitu studi literatur, perancangan sistem, flowchart dan perancangan tampilan antarmuka.

BAB 4 Hasil dan Diskusi Pada bab ini berisi mengenai hasil yang diperoleh dalam penelitian ini. Hasil ini diperoleh berdasarkan hasil dari sistem yang telah berhasil dibangun, dan hasil analisis data yang didapatkan selama pengerjaan skripsi.

BAB 5 Simpulan dan Saran Pada bab ini berisi mengenai jawaban atas tujuan penelitian yang telah dijelaskan sebelumnya, dan informasi tambahan yang diperoleh dari hasil penelitian. Bab ini juga berisi tentang saran dari penulis terkait penelitian ini.

UMMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA