

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Di Indonesia, perhatian masyarakat terhadap penyandang disabilitas masih sedikit jika dibandingkan dengan negara luar. Penyebab sedikitnya perhatian terhadap penyandang disabilitas salah satunya adalah kurangnya informasi atau pengetahuan mengenai penyandang disabilitas (ILO, 2014). Salah satu cara untuk mendapatkan informasi mengenai penyandang disabilitas adalah dengan membaca berita. Pemberitaan yang dilakukan dengan baik dapat membantu masyarakat memahami penyandang disabilitas (ILO, 2014; McPherson *et al.*, 2016; Tang and Bie, 2016; Rees, Robinson and Shields, 2019).

Oleh karena itu, untuk melakukan pemberitaan jurnalis dapat melihat berita-berita yang telah dipublikasikan, tetapi dibutuhkan waktu yang lama dan tenaga yang banyak untuk membaca berita-berita tersebut (Gunawan, Harahap and Fadillah Rahmat, 2019). Salah satu solusi terhadap masalah tersebut adalah *Text Summarization* (Maryana, Kurniawati and Agustien, 2018; Gunawan, Harahap and Fadillah Rahmat, 2019). *Text Summarization* dapat mengekstrak kalimat-kalimat penting yang ada di dalam suatu teks berita, dan meringkasnya (Hingu, Shah and Udmale, 2015; Gunawan, Harahap and Fadillah Rahmat, 2019). *Text Summarization* juga membantu pembaca untuk mengerti isi berita dan menentukan untuk membaca keseluruhan berita cukup dengan membaca ringkasannya saja (Maryana, Kurniawati and Agustien, 2018; Du and Huo, 2020).

Salah satu cara yang paling sering digunakan untuk melakukan *Text Summarization* adalah dengan TF-IDF (*Term Frequency - Inverse Document Frequency*) dan LSA (*Latent Semantic Analysis*) (Widyassari *et al.*, 2020). TF-IDF digunakan untuk menentukan *weight* dari setiap kata, sehingga dapat menemukan kalimat penting dari suatu teks (Gunawan, Juandi and Soewito, 2015). LSA digunakan karena dapat menghasilkan *weight* terhadap kalimat sesuai dengan semantiknya (makna yang terkandung di dalamnya) selayaknya manusia dengan melakukan pemetaan terhadap kalimat ke dalam ruang konsep abstrak berdimensi rendah untuk mendapatkan konsep yang diwakilkan oleh sekumpulan kata (Chowdhury, Sarkar and Dam, 2018). Selain itu, LSA mengurangi *noise* dengan menjamin relevansi terhadap dokumen asli dan tidak melebih-lebihkan hasil ringkasannya (Geetha and Deepamala, 2015). Kelebihan lain dari LSA adalah *unsupervise*, yaitu tidak diperlukan training dan pengetahuan eksternal dan dapat menunjukkan struktur kata, kalimat, atau teks yang tidak terlihat (Afsharizadeh, Ebrahimpour-Komleh and Bagheri, 2018; Lwin and Nwet, 2018).

Pada penelitian sebelumnya yang dilakukan oleh Saputra, Fachrurrozi, dan Yunita (2017), yang berjudul “Peringkasan Teks Berita Berbahasa Indonesia Menggunakan Metode *Latent Semantic Analysis* (LSA) dan Teknik Steinberger & Jezek”, menguji metode LSA dan teknik Steinberger & Jezek dengan menggunakan 10 teks berita. Akan tetapi berdasarkan penelitian yang dilakukan oleh Lwin dan Nwet (2018) menunjukkan bahwa *cross method* memiliki hasil yang lebih baik dibandingkan dengan teknik Steinberger & Jezek.

Berdasarkan penelitian yang dilakukan oleh Lwin dan Nwet (2018), ROUGE (*Recall-Oriented Understudy for Gisting Evaluation*) merupakan matriks

evaluasi yang paling sering digunakan untuk *Text Summarization*. Tetapi, berdasarkan penelitian Adelia, Suyanto, dan Wisesty (2019), penggunaan ROUGE pada Bahasa Indonesia memiliki nilai yang rendah yang disebabkan oleh perbedaan kaidah Bahasa Indonesia dan Bahasa Inggris, seperti penggunaan kata “perkembangan” pada teks sumber untuk menggambarkan “*growth*”, sedangkan pada ringkasan menggunakan kata “pertumbuhan” yang memiliki arti yang sama dengan “perkembangan”. Oleh karena itu, untuk melakukan evaluasi terhadap ringkasan yang menggunakan Bahasa Indonesia digunakan *precision*, *recall*, dan *F1-score* sebagai pengganti *ROUGE Score* (Gunawan, Juandi and Soewito, 2015; Utama, Barakbah and Helen, 2017; Sabuna and Setyohadi, 2018; Prasetyo, Kaswidjanti and Himawan, 2020).

Pada penelitian yang dilakukan oleh Steinberger dan Jezek (2009), yang berjudul “*Evaluation Measures For Text Summarization*”, membahas tentang beberapa metode evaluasi untuk *Text Summarization* salah satunya adalah dengan menggunakan *content-based measures*. *Content-based measures* digunakan karena dapat mengevaluasi hasil ringkasan walaupun ringkasan tersebut merupakan parafrase (tidak sama persis) dari teks aslinya. Dalam penelitiannya, Steinberger menggunakan nilai *cosine similarity* antara hasil ringkasan mesin dan ringkasan manusia sebagai bagian dari evaluasi *content-based measures*. Berdasarkan latar belakang yang telah dijelaskan sebelumnya, penelitian ini menggunakan LSA untuk meringkas berita mengenai penyandang disabilitas atau difabel.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan, rumusan masalah yang dipilih adalah sebagai berikut.

1. Bagaimana cara mengimplementasi metode *Latent Semantic Analysis* untuk meringkas berita tentang penyandang disabilitas?
2. Berapa nilai *cosine similarity*, *precision*, *recall*, dan *F1-score* dari metode *Latent Semantic Analysis* dalam meringkas berita tentang penyandang disabilitas?

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini dapat dijabarkan menjadi beberapa poin sebagai berikut.

1. Berita yang digunakan menggunakan Bahasa Indonesia.
2. *Dataset* yang digunakan ada dua yaitu, *dataset* IndoSum dan *dataset* yang diambil dari situs liputan6.com, newsdifabel.com, difabel.tempo.co, kompas.com, cnnindonesia.com, kumparan.com, merdeka.com, juara.bolasport.com, dan antaranews.com.
3. *Dataset* pertama berisi 66 berita difabel yang merupakan kumpulan berita dari tahun 2016 sampai dengan 2018.
4. *Dataset* kedua berisi 50 berita difabel yang merupakan kumpulan berita dari tahun 2016 sampai dengan 2021.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut.

1. Mengimplementasikan metode *Latent Semantic Analysis* dalam meringkas berita tentang penyandang disabilitas.
2. Mengukur nilai *cosine similarity*, *precision*, *recall*, dan *F1-score* dari metode *Latent Semantic Analysis* dalam meringkas berita tentang penyandang disabilitas.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah untuk membantu para jurnalis melakukan pemberitaan yang baik dengan membaca ringkasan berita-berita sebelumnya. Penelitian ini juga membantu menghemat waktu dan tenaga pembaca berita dalam memahami inti dari berita sehingga informasi mengenai penyandang disabilitas dapat diketahui oleh masyarakat luas dan meningkatkan kesadaran masyarakat terhadap penyandang disabilitas.

1.6 Sistematika Penulisan

Sistematika penulisan pada laporan skripsi ini dibagi atas 5 (lima) bab utama, yaitu sebagai berikut.

1. BAB 1 PENDAHULUAN

Bab satu menguraikan tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

2. BAB 2 LANDASAN TEORI

Bab dua menguraikan tentang teori-teori yang mendasari penelitian, yaitu *Text Summarization*, *Text Preprocessing*, Disabilitas atau Difabel, *Text Preprocessing*, *LSA*, *One-Hot Encoding*, *FastText*, *Cosine Similarity*, dan *Confusion Matrix*.

3. BAB 3 METODOLOGI PENELITIAN

Bab tiga berisikan pembahasan terhadap tahapan metode penelitian yang dilakukan, *flowchart*, dan perancangan *interface* dari aplikasi web.

4. BAB 4 HASIL DAN DISKUSI

Bab empat menguraikan tentang implementasi dan pengujian kode, hasil pengujian, dan evaluasi terhadap hasil yang telah diperoleh.

5. BAB 5 SIMPULAN DAN SARAN

Bab lima berisikan simpulan yang menjawab tujuan penelitian, dan saran terhadap pengembangan atau penelitian yang bersangkutan di kemudian hari.