

BAB I

PENDAHULUAN

1.1 Latar Belakang

Di era inovasi 4.0, kemajuan ilmu pengetahuan dan inovasi saat ini sangat pesat, khususnya inovasi jaringan web. Hampir seluruh wilayah di dunia memanfaatkan inovasi ini untuk mencari dan menyebarkan data. Demikian pula, sebagian besar masyarakat Indonesia menggunakan internet untuk mengumpulkan dan mendistribusikan informasi. Berdasarkan survei APJII 2018, Pulau Jawa dan Sumatera masih memiliki persentase pengguna internet tertinggi. Berdasarkan temuan survei, sebuah portal web berita akan dibuat yang secara eksklusif menyediakan berita terkait pertanian melalui penggunaan internet yang luas. Di sini, pertanian mengacu pada aktivitas manusia dalam menggunakan sumber daya hayati untuk menghasilkan makanan, bahan baku industri, atau sumber energi (Devianto, 2021). Portal berita *online* menjadi salah satu media massa yang memiliki kekuatan penting dalam menyebarkan sebuah informasi. Permasalahan yang sering dikeluhkan pengguna adalah bagian antarmuka portal yang sulit dimengerti, dan fungsionalitas yang belum dapat digunakan dengan maksimal (Handiyani dalam Fajarini, 2020). Menurut (Fajarini, 2020) dalam penelitiannya yang berjudul “Evaluasi Portal Berita *Online* Pada Aspek *Usability* Menggunakan *Heuristic Evaluation* dan *Think Aloud*” Portal berita *online* harus memenuhi beberapa aspek untuk bisa menjaga kualitas penggunaan dan informasi yang disediakan, salah satunya aspek *usability*. Aspek *usability* menentukan tingkat kebermanfaatan portal berita yang dinilai dari tingkat efektivitas, efisiensi, serta kepuasan dengan melakukan evaluasi *usability*.

Clickbait adalah judul konten yang dibuat untuk menarik perhatian dan mendorong pengunjung untuk mengklik tautan ke halaman web tertentu. *Clickbait* disebut juga tautan jebakan pada judul konten yang dibuat sedemikian rupa guna menarik perhatian pembaca. Namun, isi kontennya biasa saja dan terkadang tidak relevan dengan judulnya. Praktik *clickbait* banyak dijumpai dalam

berita online serta konten yang ada di media sosial (Daradinanti, 2022). Menurut (Sagita, 2020) dalam penelitiannya yang berjudul “Klasifikasi Berita *Clickbait* Menggunakan *K-Nearest Neighbor*”, Salah satu cara untuk meningkatkan *traffic* pembaca dan pengunjung adalah melalui *clickbait*. Praktik konten yang menyesatkan ini berdampak pada pemasok situs berita karena minat pengguna dan kesulitan pengguna dalam memilih konten yang menyesatkan atau konten berita yang tidak menyesatkan. Tindakan konten menyesatkan sendiri sangat bergantung pada penyedia situs berita yang sering menggunakan judul berbahaya untuk menarik pengguna. Untuk mendapatkan hasil berita konten yang menyesatkan maupun konten yang tidak menyesatkan digunakan *K-Nearest Neighbors* (K-NN) dimana waktu penanganan pada KNN lebih cepat dibandingkan dengan waktu penanganan lainnya. Dari konsekuensi pemeriksaan yang mengarah pada urutan konten berita yang menyesatkan, hasil terbaik diperoleh pada $k = 11$ melibatkan situasi 1 bagian informasi dengan jumlah 800 informasi persiapan dan 200 informasi tes yang menyampaikan akurasi 71%, presisi 72%, dan *recall* 71%. Hal ini menunjukkan bahwa urutan berita konten menyesatkan dapat dikarakterisasi menggunakan *K-Nearest Neighbor* (Sagita, 2020). Untuk menggunakan algoritma KNN sendiri penulis tentu memerlukan pembobotan dari metode TF-IDF. Menurut (Sierra, 2019) metode ini juga terkenal efisien, mudah dan memiliki hasil yang akurat. Metode ini akan menghitung nilai *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF) pada setiap token (kata) di setiap dokumen dalam korpus. Secara sederhana, metode TF-IDF digunakan untuk mengetahui berapa sering suatu kata muncul di dalam dokumen.

Dibantu oleh fakta bahwa masyarakat Indonesia pada umumnya akan lebih rentan terhadap dampak konten yang menyesatkan, yang tercermin dari artikel atau konten dengan judul konten yang menyesatkan memiliki rata-rata akses yang lebih tinggi dibandingkan artikel yang menggunakan judul berita tanpa komponen konten yang menyesatkan. memperkuat tujuan penulis di balik pembuatan ini adalah premis pemeriksaan dari “Penerapan *TF-IDF* dan *K-Nearest Neighbor*

Dalam Membangun Sistem Deteksi *Clickbait* Pada Judul Berita *Online* Bahasa Indonesia”.

1.2 Rumusan Masalah

Berdasarkan penjelasan latar belakang sebelumnya, permasalahan yang akan dibahas penulis adalah:

1. Bagaimana penerapan TF-IDF dapat membantu memahami ciri-ciri *clickbait* pada judul berita *online* dalam bahasa Indonesia?
2. Apakah penggunaan metode *k-Nearest Neighbor* dapat meningkatkan kemampuan sistem dalam mengenali *clickbait* berdasarkan representasi TF-IDF pada judul berita bahasa Indonesia?
3. Bagaimana variasi nilai parameter k pada algoritma *k-Nearest Neighbor* memengaruhi tingkat keberhasilan sistem dalam mendeteksi *clickbait* pada judul berita *online* bahasa Indonesia?

1.3 Batasan Masalah

Batasan masalah yang ditetapkan pada penelitian ini adalah:

1. Penelitian ini terbatas pada ketersediaan *dataset* judul berita *online* bahasa Indonesia yang bersifat representatif dan mencakup contoh *clickbait* dan *non-clickbait* secara seimbang.
2. Fokus pada implementasi metode TF-IDF terbatas pada pemilihan fitur-fitur atau kata-kata yang relevan dan memiliki dampak signifikan dalam mendeteksi *clickbait* pada judul berita.
3. Penelitian ini membatasi variasi parameter k pada *k-Nearest Neighbor* untuk mengevaluasi pengaruhnya terhadap akurasi sistem deteksi *clickbait* pada judul berita *online* bahasa Indonesia.
4. Evaluasi kinerja sistem deteksi *clickbait* akan difokuskan pada metrik seperti akurasi, presisi, *recall*, dan *F1-Score* untuk memberikan pemahaman yang komprehensif tentang efektivitas kombinasi TF-IDF dan *k-Nearest Neighbor*.

5. Penelitian ini akan menghasilkan *output* berupa program web untuk mendeteksi *clickbait* pada judul berita *online*.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk mengeksplorasi dan mengimplimentasikan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) dan algoritma *k-Nearest Neighbor* (k-NN) guna membangun sistem deteksi *clickbait* pada judul berita *online* berbahasa Indonesia.

1.5 Manfaat Penelitian

Adapun manfaat yang dapat diperoleh dari penelitian ini adalah sebagai berikut:

1. Penelitian ini diharapkan dapat memberikan kontribusi pada pemahaman lebih mendalam tentang ciri-ciri *clickbait* pada judul berita *online* berbahasa Indonesia, terutama dengan penerapan metode TF-IDF dan *k-Nearest Neighbor*.
2. Hasil penelitian ini diharapkan dapat membantu dalam pengembangan alat atau sistem deteksi *clickbait* yang efektif dan dapat diimplementasikan pada platform berita *online*, memberikan manfaat dalam meningkatkan keaslian dan kualitas berita.
3. Dengan adanya sistem deteksi *clickbait* yang lebih baik, diharapkan penelitian ini dapat berkontribusi pada peningkatan kualitas informasi yang disajikan kepada pembaca, membantu mengurangi penyebaran berita palsu atau sensasional.
4. Penelitian ini dapat memberikan wawasan terkait optimalisasi penggunaan metode TF-IDF dan algoritma k-NN dalam konteks deteksi *clickbait*, memandu peneliti dan praktisi untuk mengambil keputusan yang lebih tepat dalam implementasi teknik-teknik ini.

5. Temuan dari penelitian ini dapat menjadi dasar bagi penelitian selanjutnya dalam bidang deteksi *clickbait*, memberikan landasan untuk penelitian lebih lanjut terkait metode-metode baru atau perbaikan teknis dalam konteks bahasa Indonesia.