

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Penggunaan internet telah menjadi sangat populer di era digital, dengan dampak yang luas pada berbagai aspek kehidupan manusia. Internet telah menyentuh berbagai bidang seperti gaya hidup, pendidikan, penelitian, dan dunia bisnis. Salah satu keuntungan utamanya adalah kemampuannya sebagai sumber informasi yang cepat dan fleksibel. Namun, terkadang ada isu dengan informasi yang disajikan secara menarik untuk meningkatkan jumlah kunjungan, meskipun judul tidak selalu sesuai dengan isi artikel di media elektronik. Fenomena ini dikenal dengan istilah judul *trapping* atau *Clickbait*, yang menjadi salah satu strategi media online untuk mendapatkan lebih banyak traffic pengunjung (Yavi, A. F, 2018).

Penggunaan clickbait telah menjadi cara untuk meningkatkan pendapatan dengan menarik lebih banyak pembaca dan pengunjung. Semakin banyak pengunjung yang mengakses sebuah situs, semakin besar potensi untuk menghasilkan pendapatan dari situs tersebut. Tren clickbait di media online dipicu oleh persaingan yang semakin ketat antar media dalam upaya untuk menarik pembaca dan mendapatkan pendapatan dari kunjungan situs dan iklan. Meskipun clickbait memberikan keuntungan materi, terkadang judul-judul yang digunakan tidak sepenuhnya relevan dengan isi artikel atau hanya memberikan gambaran yang berlebihan atau tidak lengkap (Yamlean, Y, 2019).

Clickbait yang demikian nyatanya dapat menurunkan kepercayaan pembaca (Indrajaya & Lukitawati, 2019). Hal ini dikarenakan strategi yang bekerja dengan cara mengeksploitasi *curiosity gap* atau rasa keingintahuan manusia (Chakraborty et al., 2017) itu tidak menampilkan informasi yang merepresentasikan isi berita pada judul, bahkan kerap melebih-lebihkan berita, sehingga menciptakan kesalahpahaman yang akhirnya tidak memenuhi ekspektasi pembaca (Anand et al., 2017).

Berdasarkan praktik clickbait yang marak saat ini, penelitian ini melakukan klasifikasi berita clickbait atau non-clickbait dengan menggunakan algoritma K-Nearest Neighbor (KNN). Algoritma KNN digunakan untuk mengkategorikan berita berdasarkan kemiripannya dengan data latih yang telah diberi label. Dalam penelitian ini, klasifikasi dilakukan berdasarkan atribut pada dataset yang terdiri dari judul berita clickbait, label, dan

label-score. Dengan memanfaatkan algoritma KNN, diharapkan dapat membantu mengklasifikasikan berita dengan lebih akurat dan efisien.

Dalam penelitian ini digunakan data pendukung dari Mendeley Data (CLICK-ID: A Novel Dataset for Indonesian Clickbait Headlines) yang berisi kumpulan headline yang diambil dari 12 portal berita online lokal. Data tersebut kemudian diolah dengan menggunakan algoritma K-Nearest Neighbor.

penelitian ini menghasilkan sebuah tampilan berupa dashboard yang menampilkan informasi dari dataset yang telah dipreprocessing. Tampilan dashboard ini memberikan informasi seputar judul berita clickbait dan non-clickbait. Dengan menggunakan tampilan dashboard, pengguna dapat dengan mudah melihat dan menganalisis informasi mengenai klasifikasi berita clickbait dan non-clickbait secara visual dan interaktif. Tampilan dashboard ini diharapkan dapat membantu para peneliti atau pengguna lainnya untuk memahami pola dan karakteristik judul berita clickbait, serta memberikan wawasan yang berguna dalam memahami fenomena clickbait di media online Indonesia.

1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah dikemukakan sebelumnya, maka rumusan masalah dalam penelitian ini adalah:

1. Bagaimana mengimplementasikan metode K-Nearest Neighbor untuk mengklasifikasikan judul berita clickbait dan non-clickbait?
2. Berapa tingkat akurasi algoritma K-Nearest Neighbor dalam melakukan klasifikasi?
3. Berapa nilai K yang menghasilkan akurasi paling tinggi untuk mengklasifikasikan berita clickbait dan non-clickbait?
4. Apa saja informasi tentang berita clickbait dan non-clickbait dari dataset yang dapat ditampilkan dalam sebuah *dashboard*?

1.3 Tujuan Penelitian

1. Menerapkan algoritma K-Nearest Neighbor untuk klasifikasi berita clickbait dan non-clickbait.

2. Mengetahui tingkat akurasi dari algoritma K-Nearest Neighbor dalam melakukan klasifikasi.
3. Mengetahui nilai k yang menghasilkan nilai akurasi tertinggi untuk klasifikasi berita clickbait dan non-clickbait.
4. Menampilkan informasi tentang berita clickbait dan non-clickbait dari sebuah menggunakan *dashboard*.

1.4 Batasan Masalah

Agar pengerjaan tugas akhir ini menjadi lebih terarah dan mendapatkan hasil yang lebih spesifik, maka penulisan ini akan dibatasi dengan batasan masalah sebagai berikut:

1. Data yang digunakan adalah dari website Mendeley Data ‘CLICK-ID: A Novel Dataset for Indonesian Clickbait Headlines’.
2. Data berita yang digunakan yaitu berita pada tahun 2019 dengan total 12 portal Berita.
3. Klasifikasi terdiri dari 2 kelas yaitu clickbait dan non-clickbait.
4. Penentuan kategori berita yang digunakan adalah berita berlabel (clickbait) yang menggunakan gaya bahasa yang lebih menarik dalam judul berita.

1.5 *State Of the Art*

Judul Jurnal	Pembahasan
<p>Clickbait di Media Online Indonesia</p> <p>Peneliti Yayat D. Hidayat</p> <p>Lokasi Makassar</p> <p>Tahun 2019</p>	<p>Hasil Penelitian:</p> <p>Jurnal ini melakukan penelitian untuk mengetahui bagaimana cara media <i>online</i> mengemas judul-judul artikel yang mengandung <i>clickbait</i>.</p> <p>Alasan Menjadi Tinjauan Penelitian:</p> <p>Jurnal ini digunakan sebagai pengetahuan tentang penggunaan headline berita <i>clickbait</i> yang dilakukan oleh portal berita <i>online</i> dengan analisis data.</p>

<p>Nama Jurnal Jurnal Pekommas</p>	
<p>Analisis Perbandingan Algoritma SVM dan KNN Untuk Klasifikasi Anime Bergenre Drama</p> <p>Peneliti Vika Vitaloka Pramansah, Dadang Iskandar Mulyana, Titi Silfia</p> <p>Lokasi Universitas Janabadra</p> <p>Tahun 2022</p> <p>Nama Jurnal Jurnal Informatika dan Teknologi Informasi</p>	<p>Hasil Penelitian: Jurnal ini melakukan penelitian untuk mengekstraksi ciri dokumen dan kemiripan dokumen</p> <p>Alasan Menjadi Tinjauan Penelitian: Dengan melihat hasil algoritma KNN pada jurnal ini yang cukup baik dibanding SVM, menjadikan langkah awal dalam pendalaman algoritma KNN hingga akhirnya dijadikan algoritma dalam pengembangan sistem ini.</p>
<p>Klasifikasi Topik berita Berbahasa Indonesia Menggunakan k-Nearest Neighbor</p> <p>Peneliti Andi Ahmad Irfa, Adiwijaya, Mohamad Syahrul Mubarak</p> <p>Lokasi Universitas Telkom</p> <p>Tahun 2018</p>	<p>Hasil Penelitian: Jurnal ini melakukan penelitian untuk mengklasifikasikan berita berdasarkan topik dari berita tersebut menggunakan algoritma KNN. Menghasilkan akurasi dengan total 69.9% dengan nilai k=16.</p> <p>Alasan Menjadi Tinjauan Penelitian: Jurnal ini memiliki objek penelitian yang sama yaitu klasifikasi berita menggunakan algoritma KNN dan text mining, sehingga dapat mengetahui</p>

<p>Nama Jurnal e-Proceeding of Engineering</p>	<p>implementasi KNN dalam mengklasifikasikan berita.</p>
<p>Perbandingan Algoritma Naive Bayes dan K-Nearest Neighbor Untuk Klasifikasi Berita Hoax Kesehatan Di Media Sosial Twitter</p> <p>Peneliti Dheo Alpian Nanda Krisna, Umniy Salamah</p> <p>Lokasi Universitasi Mercu Buana</p> <p>Tahun 2022</p> <p>Nama Jurnal Jurnal Teknik Informatika Kaputama (JTIK)</p>	<p>Hasil Penelitian: Jurnal ini melakukan perbandingan algoritma KNN dan Naive Bayes dalam klasifikasi berita hoax di media sosial Twitter. Hasil dari penelitian ini menunjukkan algoritma KNN memberikan akurasi lebih baik 68% daripada algoritma Naive Bayes 66%.</p> <p>Alasan Menjadi Tinjauan Penelitian: Digunakannya jurnal ini untuk mendapatkan wawasan lebih dalam tentang penerapan algoritma K-Nearest Neighbors (KNN) dalam klasifikasi berita. Hal ini akan membantu memperkaya pengetahuan mengenai penerapan metode KNN pada data dinamis seperti berita online dan memperluas wawasan tentang teknik penggunaan TF-IDF.</p>
<p>Penerapan Metode Euclidian Distance Untuk Ekstraksi Ciri Dokumen dan Kemiripan Dokumen</p> <p>Peneliti Yunitasari</p>	<p>Hasil Penelitian: Jurnal ini melakukan penelitian untuk mengekstraksi ciri dokumen dan kemiripan dokumen dengan menggunakan metode TF-IDF dan <i>euclidian distance</i> untuk melakukan ekstraksi ciri pada dokumen.</p>

<p>Lokasi Universitas PGRI Madiun</p> <p>Tahun 2019</p> <p>Nama Jurnal Double Click: Journal of Computer and Information Technology</p>	<p>Alasan Menjadi Tinjauan Penelitian: Jurnal ini digunakan sebagai pengetahuan mengenai ekstraksi ciri menggunakan <i>TF-IDF</i> untuk mendapatkan fitur-fitur yang terdapat dalam dokumen untuk proses <i>text mining</i>.</p>
--	---

1.6 Sistematika Penulisan

Penyusunan Tugas Akhir ini diuraikan dalam beberapa pokok permasalahan yang terbagi dalam beberapa bab agar dapat memberikan gambaran secara umum mengenai permasalahan dan pemecahannya. Sistematika penulisan Tugas Akhir ini adalah sebagai berikut:

Bab 1 Pendahuluan

Bab ini memuat pendahuluan penelitian yang terdiri dari latar belakang, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah penelitian, *state of the art*, serta sistematika penulisan penelitian.

Bab 2 Landasan Teori

Bab ini memuat landasan teori penelitian yang terdiri dari teori dasar mengenai *Clickbait*, *Text Mining*, *NLP*, ekstraksi fitur (*TF-IDF*), Algoritma *k-Nearest Neighbor*, dan evaluasi (*confusion matrix*).

Bab 3 Analisis dan Perancangan

Bab ini memuat tentang analisis dan perancangan mengenai Perancangan sistem dan perancangan fungsi.

Bab 4 Implementasi dan Pengujian

Bab ini memuat tentang implementasi program, dan proses pengujian sistem.

Bab 5 Kesimpulan dan Saran

Bab ini memuat kesimpulan dan saran dari pembahasan dan pengujian pada sistem.

Daftar Referensi**Lampiran**