

BAB 1

PENDAHULUAN

1.1 LATAR BELAKANG

Berdasarkan hasil survei Asosiasi Penyelenggara Jasa Internet Indonesia (APJII), pengguna internet di Indonesia mencapai 215,63 juta orang pada periode 2022-2023. Angka ini menunjukkan peningkatan sebesar 2,67% dibandingkan periode sebelumnya yang mencatatkan 210,03 juta pengguna. Jumlah tersebut setara dengan 78,19% dari total penduduk Indonesia yang mencapai 275,77 juta jiwa, meningkat sebesar 1,17% dari tingkat penetrasi internet pada 2021-2022 yang tercatat sebesar 77,02%. Berdasarkan analisis lebih lanjut, tingkat penetrasi internet menunjukkan perbedaan berdasarkan jenis kelamin. Pengguna internet laki-laki mencapai tingkat penetrasi 79,32%, sedikit lebih tinggi dibandingkan dengan perempuan yang berada pada angka 77,36%. Selain itu, perbedaan penetrasi juga terlihat berdasarkan lokasi geografis. Di kawasan perkotaan, tingkat penetrasi internet pada periode 2022-2023 tercatat sebesar 77,36%, sedangkan di kawasan perdesaan tingkat penetrasi bahkan lebih tinggi, yaitu 79,79%. Data ini mencerminkan dinamika penggunaan internet di Indonesia yang terus berkembang, baik secara keseluruhan maupun dalam konteks demografis tertentu, menunjukkan potensi besar bagi pengembangan teknologi digital dan infrastruktur komunikasi di berbagai wilayah [1].

Berita merupakan sumber informasi penting yang menyajikan peristiwa terkini dan dapat ditemukan di berbagai media massa seperti surat kabar, televisi, serta platform lainnya. Dalam perkembangannya, kemajuan teknologi informasi telah secara signifikan mempermudah proses penyebaran berita melalui *media online*. *Media online* adalah jenis media massa yang menggunakan internet sebagai basisnya, memungkinkan akses informasi yang nyaman dan *real-time*. Pembaca kini tidak lagi perlu menunggu hingga keesokan hari untuk mendapatkan kabar terbaru; cukup dengan memanfaatkan internet, berita dapat diakses dengan cepat kapan saja dan di mana saja.

Namun, meskipun kemudahan akses ini sangat membantu, banyaknya informasi yang tersedia secara daring belum sepenuhnya diimbangi dengan kualitas

dan objektivitas dari informasi itu sendiri. Situasi ini menimbulkan tantangan baru bagi pembaca, yaitu memilah dan menyaring informasi yang relevan serta dapat dipercaya di tengah arus data yang melimpah. Proses ini tidak hanya memakan waktu tetapi juga mengharuskan pembaca memiliki kemampuan literasi informasi yang baik. Oleh karena itu, diperlukan sistem yang mampu secara efektif memilah atau mengklasifikasikan informasi berdasarkan tingkat objektivitas dan relevansinya, sehingga memudahkan pembaca untuk memperoleh informasi yang berkualitas [2].

Di era digital, kecepatan dan akurasi informasi menjadi kunci utama dalam mendukung pengambilan keputusan, baik secara individu maupun kolektif. Media online kini telah menjadi sumber utama bagi masyarakat untuk mendapatkan berita terkini. Namun, muncul tantangan signifikan dalam pengelolaan dan klasifikasi berita daring, terutama terkait ketidakseimbangan jumlah berita di antara berbagai kategori. Kondisi ini dapat menyebabkan model klasifikasi yang digunakan dalam pengolahan berita cenderung mengabaikan kategori atau kelas berita yang jumlahnya lebih sedikit (kelas minoritas). Akibatnya, hasil prediksi dari model tersebut menjadi kurang akurat dan tidak seimbang, yang pada akhirnya mengurangi kepercayaan terhadap sistem klasifikasi itu sendiri [3].

Untuk mengatasi permasalahan ini, dibutuhkan pendekatan khusus yang mampu mengatasi ketidakseimbangan kelas dalam berita daring. Pendekatan tersebut harus mampu mempertimbangkan keberagaman kategori berita secara proporsional, sehingga dapat menghasilkan model klasifikasi yang lebih akurat, andal, dan memberikan manfaat optimal bagi pengguna.

Untuk mengatasi ketidakseimbangan kelas dalam klasifikasi informasi online, Penulis menggunakan *Synthetic Minority Oversampling Technique* (SMOTE). SMOTE bekerja dengan menghasilkan sampel sintetik dari kelas minoritas, sehingga dapat meningkatkan keterwakilan dan kemampuan prediksi model untuk kelas minoritas [4]. Selain itu, Penulis juga menggunakan algoritma *K-Nearest Neighbors* (KNN). KNN menggunakan informasi dari tetangga terdekat untuk membuat prediksi, sehingga memungkinkannya berkontribusi aktif dalam pengelolaan ketidakseimbangan kelas. Penggabungan antara SMOTE dan KNN

dapat menciptakan model klasifikasi berita online yang lebih andal, mampu membuat prediksi berimbang di seluruh kategori berita.

Berdasarkan uraian latar belakang di atas, maka judul penelitian ini adalah “Implementasi SMOTE untuk Mengatasi *Imbalance data set Class* pada Berita *Online* dengan Menggunakan KNN”. Penelitian ini bertujuan untuk mengeksplorasi dan mengimplementasikan SMOTE sebagai metode resampling pada *Dataset* berita online untuk meningkatkan kinerja model klasifikasi. Selain itu, penerapan KNN sebagai algoritma klasifikasi akan dieksplorasi untuk melihat bagaimana kombinasi kedua metode ini dapat memberikan dampak positif dalam mengelola ketidakseimbangan antar kelas dalam konteks informasi berita online.

1.2 RUMUSAN MASALAH

1. Bagaimana cara menerapkan SMOTE guna menangani *imbalance data* pada data set komentar berita *online*?
2. Bagaimana cara mengevaluasi performa model klasifikasi KNN setelah penerapan SMOTE?
3. Sejauh mana SMOTE efektif dalam meningkatkan performa model KNN?

1.3 BATASAN MASALAH

Batasan masalah dari penelitian ini adalah :

1. Penelitian ini menggunakan data berita online dari sumber tertentu.
2. Fokus utama adalah penanganan ketidakseimbangan kelas pada *Dataset* berita online khususnya pada kelas minoritas.
3. KNN dipilih sebagai algoritma utama. model lain tidak akan dieksplor dlm penelitian ini.
4. Penilaian kinerja model hanya akan berfokus pada metrik klasifikasi seperti akurasi, presisi, recall dan *f1 score*.

1.4 TUJUAN

1. Dapat menerapkan SMOTE guna menangani *imbalance class* pada *Dataset* berita *online*.
2. Dapat mengevaluasi performa model klasifikasi KNN setelah penetapan SMOTE.

3. Mengetahui sejauh mana SMOTE efektif dalam meningkatkan performa model.

1.5 MANFAAT

Penelitian ini diharapkan dapat membantu meningkatkan akurasi klasifikasi dalam analisis komentar berita online dengan menggunakan metode KNN. Melalui penerapan teknik SMOTE, ketidakseimbangan kelas pada data komentar berita dapat diatasi, sehingga model klasifikasi mampu mengenali dan menangani kelas minoritas dengan lebih baik. Selain itu, hasil penelitian ini juga diharapkan dapat menjadi landasan untuk pengembangan lebih lanjut dalam teknik penanganan ketidakseimbangan data yang lebih canggih, tidak hanya pada klasifikasi komentar berita, tetapi juga pada aplikasi lain yang menghadapi tantangan serupa.

1.6 SISTEMATIKA PENULISAN

Penelitian ini disusun dalam beberapa bab yang saling terkait. Bab 1 membahas latar belakang, rumusan masalah, manfaat dan tujuan penelitian, batasan masalah, serta sistematika penulisan. Bab 2 berisi kajian pustaka yang mencakup dasar teori dan penelitian-penelitian relevan terkait penerapan SMOTE untuk menangani *imbalance class* pada klasifikasi komentar berita menggunakan metode KNN. Bab 3 menjelaskan metode penelitian, termasuk alat dan perangkat yang digunakan, alur penelitian, serta proses pengumpulan dan pengolahan data. Bab 4 memaparkan hasil penelitian, mencakup analisis penerapan SMOTE, evaluasi performa model KNN, dan interpretasi hasil klasifikasi. Akhirnya, Bab 5 menyajikan kesimpulan dari penelitian ini dan memberikan rekomendasi untuk pengembangan penelitian di masa depan.