

BAB I

PENDAHULUAN

1.1. Latar Belakang

Perkembangan teknologi informasi yang pesat telah membawa dampak signifikan pada kehidupan manusia, termasuk dalam cara berkomunikasi dan berbagi informasi. Di Indonesia sendiri jumlah pengguna internet berkembang dari tahun ke tahun (Kusnandar, 2019). Media sosial, sebagai salah satu produk dari teknologi ini, telah menjadi platform utama bagi individu untuk mengekspresikan opini, berbagi pengalaman, dan menjalin interaksi sosial. Sosial media bisa menjadi keuntungan ataupun kerugian, itu bergantung dari penggunaannya sendiri (L. Minocha, 2022). Dengan miliaran pengguna aktif setiap harinya, media sosial menghasilkan data dalam jumlah besar yang dikenal sebagai *big data*. Data ini mencakup berbagai format, seperti teks, gambar, dan video, yang dapat memberikan wawasan berharga jika diolah dengan benar.

Di sisi lain, *machine learning* telah berkembang menjadi salah satu teknologi utama dalam memanfaatkan data yang terus bertambah secara eksponensial. *Machine learning* adalah cabang kecerdasan buatan yang memungkinkan komputer untuk belajar dari data tanpa harus diprogram secara eksplisit. Penggunaan teknik *Machine Learning* (ML) telah memberikan kontribusi besar pada klasifikasi pada berbagai bidang (Briouza, Gritli, Khraief, Belghith, & Singh, 2022). Algoritma seperti *Weighted K-Nearest Neighbor (WKNN)* telah terbukti efektif dalam menyelesaikan berbagai masalah klasifikasi, termasuk dalam domain teks. Perkembangan algoritma ini memungkinkan pengolahan data yang lebih akurat, cepat, dan efisien, sehingga menjadi dasar dalam analisis berbasis data, termasuk analisis sentimen.

Analisis sentimen adalah proses mengidentifikasi dan mengekstrak opini atau emosi dari teks. Dalam konteks media sosial, analisis ini berguna untuk memahami persepsi masyarakat terhadap berbagai isu, produk, atau layanan. Tantangan utama dalam analisis sentimen adalah keberagaman gaya bahasa, penggunaan istilah

slang, dan emosi yang kompleks dalam teks. Ketika seseorang sedang mengungkapkan sesuatu yang sulit atau bermasalah dengan pengalaman yang sudah dijalani saat ini, terdapat beberapa perspektif yang itu nantinya bisa diungkapkan ke dalam kata-kata (Ana, et al., 2016). Oleh karena itu, penelitian untuk meningkatkan performansi algoritma seperti WKNN menjadi penting, khususnya dalam memilih *distance metrics* yang sesuai untuk menangkap pola dalam data sentimen secara optimal.

Penelitian yang dilakukan oleh Muhammad Ezeywan Mohd Samad terkait dengan analisis perhitungan jarak pada metode KNN (Samad, Ali, & Tajuddin, 2017). Penelitian ini dilakukan untuk mengklasifikasikan diferensiasi pada senyawa yang terdapat pada minyak gaharu. Eksperimen ini menggunakan berbagai jenis *metrics* seperti *euclidean*, *city-block*, *cosine*, dan *correlation*. Dataset yang digunakan menggunakan ratio 80%-20% yang terdiri dari data *training* dan data *testing*. Penggunaan jumlah tetangga (nilai K) yang diteliti adalah sebesar 1-5. Penelitian ini memperoleh hasil penggunaan KNN dengan *metrics euclidean* dan *city-block* memperoleh akurasi 100%, diikuti dengan 89,5% diperoleh *cosine* dan *correlation*.

Pada tahun 2016, Sih-Huei Chen melakukan suatu penelitian dengan menggunakan *gaussian kernel* yang digunakan untuk mengklasifikasikan *speech emotion* dalam bentuk suara (Sih-Huei Chen, et al., 2016). Penelitian ini melakukan proses ekstraksi fitur suara terlebih dahulu dengan *nonnegative matrix factorization* (NMF) Kemudian selanjutnya dilakukan proses klasifikasi menggunakan *gaussian proses*. Hasil akurasi yang didapatkan pada penelitian ini adalah sebesar 77.74%.

Pada penelitian yang dilakukan oleh Ishika Saini dan Vikas Maheshkar membahas mengenai apakah ada pengaruh perhitungan jarak pada deteksi warna pada warna histogram dengan menggunakan KNN (Saini & Maheshkar, 2022). Perhitungan jarak yang dipakai pada penelitian ini adalah *Euclidean*, *Chi Square*, *Chebyshev* dan *Hamming*. Hasil yang didapat pada eksperimen tersebut adalah bahwasannya perhitungan dengan menggunakan *Euclidean distance* ini memiliki performa yang lebih baik dibandingkan ketiga metode perhitungan lainnya.

Penelitian yang dilakukan oleh Yujie Song dan Fei Pei juga berisi tentang bagaimana variasi *distance metrics* ini berpengaruh pada hasil akhir efektivitas pada suatu metode (Song & Pei, 2023). Metriks yang digunakan ada 5 macam, yaitu *Euclidean*, *Cosine Similarity*, *Correlation*, *Spearman* dan *Jaccard*. Hasil yang diperoleh pada penelitian menunjukkan bahwa pada dataset UCI yang digunakan didapatkan hasil perhitungan *Jaccard* memiliki rata rata akurasi yang paling tinggi dibandingkan dengan perhitungan *metrics* lainnya.

Salah satu kelemahan dari metode *K-Nearest Neighbor* (KNN) adalah metode ini memperhitungkan jarak antar tetangga sesuai dengan jumlah K yang ditentukan, Kemudian memasukkan data *test* ke dalam kategori yang terbanyak. Namun metode ini mengabaikan bobot dan kontribusi yang dimiliki oleh setiap tetangganya. Oleh karena ini pada penelitian kali ini penulis ingin menambahkan bobot dengan menggunakan perhitungan *Gaussian Kernel* untuk menghitung bobot masing-masing dari tetangga yang dipilih sesuai dengan nilai K yang dinamakan dengan *Weighted K-Nearest Neighbor* (WKNN). Di sisi lain metode KNN ini memiliki perhitungan jarak metriks yang cukup beragam, mulai dari *Euclidean* yang paling populer digunakan, Kemudian ada *Minkowski*, *Cosine Similarity*, dan *Jaccard* yang masing-masing memiliki perhitungan yang berbeda. Berangkat pada persoalan diatas, penulis ingin melakukan penelitian dengan mengkomparasikan beberapa varian perhitungan metriks pada WKNN untuk digunakan pada analisis sentimen di sosial media.

SEKOLAH PASCASARJANA

1.2. Tujuan Penelitian

Tujuan dilakukan penelitian ini adalah sebagai berikut:

1. Merancang dan mengimplementasikan perhitungan bobot pada metode KNN menggunakan *Gaussian Kernel* sehingga menjadi *Weighted K-Nearest Neighbor* (WKNN).
2. Mengembangkan sistem analisis performa WKNN dengan melakukan perhitungan dan komparasi dari berbagai *distance metrics* (*Euclidean*, *Minkowski*, *Cosine Similarity*, dan *Jaccard*) untuk menghitung tingkat akurasi, presisi, dan *recall* pada analisis sentimen di media sosial.

3. Mengevaluasi adakah pengaruh *distance metrics* (*Euclidean*, *Minkowski*, *Cosine Similarity* dan *Jaccard*) terhadap performansi metode *Weighted K-Nearest Neighbor* (*WKNN*)

1.3. Manfaat Penelitian

Manfaat dari penelitian yang dilakukan adalah sebagai berikut :

1. Memberikan kontribusi dalam pengembangan ilmu pengetahuan, khususnya dalam analisis sentimen, dengan mengevaluasi efektivitas masing-masing *distance metrics* (*Euclidean*, *Minkowski*, *Cosine Similarity*, dan *Jaccard*) pada metode WKNN untuk meningkatkan nilai akurasi, presisi, dan recall.
2. Menyediakan panduan dan rekomendasi bagi peneliti atau praktisi teknologi informasi tentang pengaruh *distance metrics* terhadap performansi WKNN, sehingga dapat diterapkan pada sistem serupa di masa depan.



SEKOLAH PASCASARJANA