

BAB I

PENDAHULUAN

I.1. Latar Belakang

Pesatnya perkembangan teknologi kecerdasan buatan maka berkembanglah metode *machine learning*, yaitu mesin yang dikembangkan untuk bisa belajar dengan sendirinya tanpa arahan dari penggunanya dimana dalam pengembangannya berdasarkan disiplin ilmu lain seperti statistika, matematika dan data mining. Metode klasifikasi pada *machine learning* yang sering digunakan pada metode *machine learning* adalah *Classification and Regression Trees (CART)*, *Random Forest*, *Naïve Bayes*, *Support Vector Machines (SVM)*, dan lain-lain. (Istiqomatul & Pardomuan 2021)

Klasifikasi merupakan bagian dari supervised learning, dimana pengkategorian data sangat membutuhkan variabel-variabel target dalam prosesnya (Wahyono et al., 2020). Pada klasifikasi terdapat model yang paling populer diantaranya *decision/classification trees*, *Bayesian Classifiers/Naïve Bayes Classifiers*, *Neural Networks*, *Statistical Analysis*, *Genetic Algorithms*, *Rough Sets*, *K-Nearest Neighbor Classifier*, *Rule based Methods*, *Memory Based Reasoning*, *Support Vector Machines* (Rahmadi, Kaurie, & Susanti, 2020).

Algoritma *K-Nearest Neighbor (KNN)* merupakan algoritma klasifikasi jenis *supervised learning* yang mengklasifikasikan data dengan cara mengelompokkan data latih ke dalam kelompok-kelompok yang memiliki jarak terdekat dengan kelas target (Mardhyath et al., 2020). KNN merupakan metode yang paling banyak digunakan di dalam *data mining* dan *machine learning* karena implementasinya yang sederhana dan memberikan *performance* yang baik (Zhang et al., 2017). KNN terpilih sebagai salah satu

dari 10 algoritma terbaik di dalam *data mining* dan *machine learning* (Wu et al., 2008). Prinsip dari KNN cukup sederhana, dimana dimulai dengan memilih jumlah K *neighbor*, kemudian lakukan perhitungan jarak, dan ambil K terdekat berdasarkan hasil perhitungan jarak. Kemudian, di antara K terdekat ini, hitunglah jumlah titik data dalam setiap kategori dan tempatkan titik data baru ke dalam kategori yang jumlah tetangganya maksimum (Maillo et al., 2017).

Terdapat 2(dua) isu utama di dalam KNN, yaitu terkait penentuan nilai K yang ideal dan metode perhitungan *distance* (Ertuğrul & Tağluk, 2017). Penentuan nilai K sangat sensitif di dalam KNN. Nilai K yang terlalu kecil menyebabkan terlalu sensitif terhadap *noise* yang dapat berdampak pada terjadinya *overfitting* dan nilai K yang terlalu besar dapat menyebabkan terjadinya *underfitting* (Beyer et al., 1999)(Subasi & Qureshi, 2023). Beberapa penelitian sudah membuktikan hal ini, dimana memiliki jumlah terdekat yang berbeda, diperoleh kinerja yang berbeda pula baik itu dari segi akurasi, *purity* maupun jumlah iterasi dari proses klasifikasi (Hidayati & Hermawan, 2021; Rao et al., 2022; Tulus & Situmorang, 2020).

Hal lain yang berpengaruh pada KNN selain jumlah k tetangga terdekat adalah metode untuk menghitung jarak terdekat *cluster*, yang juga dapat mempengaruhi kinerja algoritma K-NN. Secara *default* KNN memang menggunakan metode perhitungan *distance* dengan menggunakan *Euclidean Distance* (Santos et al., 2020). Metode perhitungan *distance* lain yang sering digunakan adalah *Manhattan Distance* (Han et al., 2012). Beberapa penelitian yang menggunakan metode *Euclidean* dan *Manhattan* menunjukkan adanya perubahan akurasi yang dihasilkan algoritma KNN bila dikaitkan dengan perhitungan *distance* pada proses klasifikasi (Hidayi & Hermawan, 2021; Iswanto et al., 2021; Margolang et al., 2022 Wahyono et al., 2020).

Mangga adalah salah satu jenis buah yang unggul dan sangat digemari untuk dikonsumsi oleh masyarakat, buah mangga sendiri banyak digemari masyarakat karena rasanya bervariasi dengan banyaknya jenis buah mangga tersebut bagi orang awam akan merasa kesulitan dalam klasifikasi jenis-jenis buah mangga. Mangga yang sering kali dikonsumsi oleh kebanyakan masyarakat yaitu jenis mangga harum manis, mangga manalagi dan mangga apel maka daripada itu dibutuhkanlah suatu metode yang cocok untuk menyelesaikan permasalahan penelitian ini.

(Rodrigues, 2018) menunjukkan bahwa nilai k dan perhitungan *distance* yang ideal dapat memiliki kombinasi tertentu yang unik dan dapat menentukan *performance* yang ideal pada proses klasifikasi. Oleh karena itu, sulit dikatakan mana metode perhitungan *distance* yang terbaik dan nilai k yang ideal karena hal ini sangat bergantung pada permasalahan yang dihadapi.

Pada penelitian ini, metode *K-Nearest Neighbor* (KNN) digunakan untuk klasifikasi jenis buah mangga dengan cara mengelompokkan data latih ke dalam kelompok-kelompok yang memiliki jarak terdekat, pengambilan citra dilakukan menggunakan kamera ponsel. Citra yang diambil akan dijadikan sebagai data latih 40 citra buah mangga apel, 40 citra buah mangga harum, 40 citra buah mangga manalagi dan total data latih 120 data citra jenis buah mangga berikutnya data uji 10 citra buah mangga apel, 10 citra buah mangga harum manis, 10 citra buah mangga manalagi dan total data uji 30 data citra jenis buah mangga.

Metode KNN dipilih karena metode ini melakukan proses secara supervised (terawasi) artinya ada pelatihan yang dilakukan untuk masuk ke dalam kelompok mana data yang akan diklasifikasikan. Prinsip kerja KNN adalah mencari jarak terdekat antara

data yang akan diuji dengan K tetangga (neighbor) terdekatnya dalam data latih. Pada algoritma ini, nilai K (jumlah tetangga terdekat) di asumsikan nilai $K=3$ dengan cara menggunakan metode pengukuran jarak *eulidean distance* dan *manhattan distance* untuk mengetahui hasil akhir klasifikasi dan tingkat akurasi yang optimalkan pada KNN. Berdasarkan referensi penelitian komparasi kedua pengukuran pada klasifikasi deteksi penyakit stroke, digunakan variasi jumlah cluster yang berbeda untuk menemukan model dengan akurasi prediksi yang paling tinggi (Iswanto et al., 2021).

Berdasarkan permasalahan penelitian diatas, penulis mengumpulkan data berupa citra digital dan hasil penelitian sebelumnya untuk kemudian dituangkan dalam bentuk proposal tesis yang berjudul “**Klasifikasi Jenis Buah Mangga Dengan Menggunakan Metode K-NN**”.

I.2. Rumusan Masalah

Adapun rumusan masalah pada penelitian ini, berdasarkan latar belakang masalah diatas, adalah:

1. Bagaimana melakukan klasifikasi jenis buah mangga menggunakan metode K-Nearest Neighbor?
2. Bagaimana pengaruh penentuan nilai K dan pengukuran jarak *Eulidean Distance* dan *Manhattan distance* dalam klasifikasi jenis buah mangga yang dibangun menggunakan metode k-Nearest Neighbors?

I.3. Tujuan

Adapun tujuan dari penelitian berdasarkan latar belakang masalah di atas adalah sebagai berikut:

1. Melakukan klasifikasi jenis buah menggunakan metode K-Nearest Neighbor.
2. Memperoleh gambaran kinerja K-Nearest Neighbor berdasarkan penentuan nilai K dan perhitungan jarak Euclidean Distance dan Manhattan Distance.

I.4. Manfaat

Adapun manfaat yang diharapkan dari penelitian berdasarkan tujuan penelitian di atas adalah sebagai berikut:

1. Memperoleh hasil analisa pengaruh metode pengukuran jarak euclidean distance dan manhattan distance jenis buah mangga pada kinerja metode K-Nearest Neighbors dalam model yang dibangun.
2. Menghasilkan klasifikasi jenis buah mangga dengan hasil kinerjanya dalam bentuk klasifikasi dan tingkat akurasi.

I.5. Batasan Masalah

Agar penelitian ini tidak melebar terlalu jauh pembahasanya, ada beberapa Batasan masalah yang penulis berikan dalam penelitian ini. Adapun batasan masalah tersebut antara lain:

1. Dataset yang digunakan dalam penelitian ini ialah 40 data citra masing-masing buah mangga yang terdiri dari mangga apel, mangga harum manis dan mangga manalagi yang kebanyakan digemari oleh masyarakat saja.
2. Metode pengukuran jarak pada metode K-Nearest Neighbors yang digunakan adalah euclidean distance dan manhattan distance.
3. Jumlah k pada metode K-Nearest Neighbors diasumsikan nilai $k = 3$ dan di dalam penelitian ini akan digunakan nilai K sebesar 1k, 2k, dan 3 k.

4. Pengukuran kinerja didasarkan pada persentase tingkat akurasi, presisi dan recall.

I.6. Sistematika Pembahasan

Adapun sistematika pembahasan yang digunakan dalam penelitian ini adalah sebagai berikut:

BAB I : PENDAHULUAN

Pada bab ini menerangkan tentang latar belakang, rumusan masalah, tujuan, manfaat serta batasan masalah penelitian.

BAB II : TINJAUAN PUSTAKA

Bab ini berikan tentang teori-teori yang berkaitan langsung dengan penelitian seperti teori mengenai program studi, machine learning, K-Neareast Neighbors dari literatur yang diperoleh melalui buku-buku, jurnal dan penelitian penelitian sebelumnya sebagai rujukan dalam penelitian ini.

BAB III : METODOLOGI PENELITIAN

Bab ini membahas tentang bentuk penerapan metode K-Nearest Neighbor ke dalam sebuah sistem model klasifikasi jenis buah mangga dan bentuk analisa yang digunakan untuk membandingkan tingkat akurasi pengukuran jarak euclidean distance dan manhattan distance dan kombinasi nilai k.

BAB IV : HASIL

Pada bab ini membahas tentang hasil dari penelitian terhadap penggunaan metode/teknik penelitian secara terurut dan logis dan menyajikan data dalam bentuk tabel atau gambar.

BAB V : PEMBAHASAN

Pada bab ini membahas tentang pembahasan hasil dari perhitungan manual yang telah dihitung pada bab sebelumnya untuk dilakukan perhitungan dengan menggunakan Tools sehingga menghasilkan sebuah laporan simulasi.

BAB VI : KESIMPULAN DAN SARAN

Pada bab ini merupakan penguraian dari hasil penelitian yang telah dilakukan serta saran-saran yang bersifat membangun untuk kepentingan dan pengembangan.