BAB II

TINJAUAN LITERATUR

2.1 Tinjauan Pustaka

Kajian mendalam mengenai perbandingan kedua algoritma ini sangat penting untuk penentuan metode terbaik. Studi komparasi oleh Loka dan Marsal (2023) dalam konteks klasifikasi status gizi menunjukkan bahwa performa K-NN dan Naive Bayes sangat bergantung pada distribusi dan karakteristik *dataset*. Aspek komputasi juga menjadi pembanding; Sahar (2020) menemukan bahwa meskipun Naive Bayes unggul dalam kecepatan proses pada *dataset* penyakit jantung, K-NN mampu memberikan akurasi yang lebih tinggi dalam kondisi tertentu. Temuan ini memberikan landasan untuk menganalisis hasil penelitian ini, yaitu apakah akurasi harus dikorbankan demi kecepatan atau sebaliknya

Dalam konteks spesifik komoditas kopi, kedua algoritma telah memiliki preseden penerapan yang kuat. Di satu sisi, penelitian oleh Fata dan Avianto (2024) berhasil menerapkan **Naive Bayes** untuk klasifikasi mutu biji kopi Robusta, yang membuktikan kelayakan metode probabilistik untuk mengenali ciri visual. Di sisi lain, relevansi **K-NN** dalam klasifikasi biji kopi diperkuat oleh studi Fadjeri (2023) yang menggunakan K-NN untuk klasifikasi bentuk biji kopi, menunjukkan bahwa fitur morfologis sangat penting dalam proses pengenalan.

Berdasarkan kelima referensi ilmiah (Budianita et al., 2024; Pemungkas dan Asnawi, 2024; Loka dan Marsal, 2023; Sahar, 2020; Fadjeri, 2023) yang semuanya diterbitkan dalam kurun waktu lima tahun terakhir, penelitian ini mengambil posisi untuk melakukan komparasi langsung dan empiris. Tujuannya adalah untuk menentukan secara objektif algoritma mana (K-NN atau Naive Bayes) yang memberikan tingkat akurasi

tertinggi untuk klasifikasi jenis biji kopi (Arabika vs. Robusta) berdasarkan ekstraksi fitur citra digital.

Perbandingan K-NN dan Naive Bayes dalam Berbagai Studi.

Perbandingan kedua algoritma ini dalam berbagai penelitian menunjukkan bahwa performa optimalnya bergantung pada karakteristik data yang digunakan:

1. **Klasifikasi Status Gizi Balita:** Performa kedua algoritma tergantung pada jumlah data dan distribusi fitur (Loka dan Marsal, 2023).
2. **Klasifikasi Tingkat Keparahan Korban Kecelakaan Lalu Lintas:** K-NN menghasilkan akurasi yang lebih tinggi pada data dengan pola spasial yang kuat, sedangkan Naive Bayes unggul pada data dengan distribusi probabilistik yang jelas (Wisdayani, 2019).
3. **Dataset Penyakit Jantung:** Naive Bayes lebih cepat dalam proses komputasi, namun K-NN lebih akurat dalam beberapa kondisi (Sahar, 2020).
4. **Klasifikasi Keluarga Miskin:** Pemilihan fitur sangat mempengaruhi akurasi klasifikasi, dan pada kondisi tertentu, Naive Bayes memiliki hasil yang lebih stabil dibandingkan K-NN (Pemungkas dan Asnawi, 2024).
5. **Klasifikasi Biji Kopi** menggunakan K-NN. Jurnal ini menegaskan relevansi K-NN untuk klasifikasi biji kopi berdasarkan fitur visual, melengkapi studi Fata dan Avianto, (2024) yang menggunakan Naive Bayes.
6. **Studi Kasus Klasifikasi Kopi Menggunakan Metode KNN dan CNN (Dengan Perhitungan Manual):** Han, Kamber, dan Pei (2012) menjelaskan konsep dasar klasifikasi data mining serta pentingnya pemilihan fitur dan algoritma yang sesuai dengan karakteristik data, yang menjadi landasan teori pada penelitian ini. Goodfellow, Bengio, dan Courville (2016) menunjukkan bahwa

CNN sangat efektif untuk pengenalan citra, namun membutuhkan data dan komputasi yang besar sehingga kurang efisien untuk dataset menengah. Witten et al. (2017) menegaskan bahwa KNN bekerja optimal pada data dengan kedekatan fitur yang kuat, sedangkan Naïve Bayes bergantung pada asumsi independensi fitur, yang sejalan dengan hasil penelitian ini. Putra dan Santoso (2021) membuktikan efektivitas KNN dalam klasifikasi biji kopi Arabika dan Robusta berbasis citra. Sari dan Nugroho (2022) menunjukkan bahwa CNN menghasilkan akurasi tinggi dalam pengenalan biji kopi, namun dengan kompleksitas sistem yang lebih tinggi. Berdasarkan studi tersebut, penelitian ini berfokus pada perbandingan KNN dan Naïve Bayes sebagai metode yang lebih sederhana dan efisien untuk klasifikasi biji kopi berbasis citra digital.

2.2 Landasan Teori

2.2.1 Deskripsi Umum Kopi

Kopi (*Coffea*) adalah salah satu komoditas pertanian terpenting di dunia yang tumbuh di kawasan tropis dan subtropis. Dari ratusan spesies kopi yang ada, dua jenis utama yang mendominasi perdagangan global dan menjadi fokus dalam penelitian ini adalah **Kopi Arabika (*Coffea arabica*)** dan **Kopi Robusta (*Coffea canephora*)**. Kedua jenis ini memiliki perbedaan signifikan yang menjadi dasar klasifikasi, baik secara manual maupun otomatis (FTNews, 2023).

2.2.2 Pengolahan Citra Digital dan K-NN

Pengolahan Citra Digital adalah serangkaian teknik pemrosesan citra menggunakan perangkat lunak dan algoritma untuk memperoleh informasi dari objek visual (Bachtiar, 2021). Dalam konteks klasifikasi biji kopi, citra yang ditangkap melalui kamera diproses untuk mengekstrak ciri-ciri fisik (fitur) yang membedakan jenis Arabika dan Robusta. Salah satu pendekatan krusial dalam pengolahan citra adalah **Ekstraksi Fitur**, yaitu proses mengubah data piksel mentah menjadi sekumpulan nilai numerik yang ringkas dan informatif. Fitur yang diekstrak meliputi **warna, tekstur**, dan **bentuk (morfologis)** biji, yang akan digunakan sebagai *input* pada algoritma klasifikasi.

Fitur-fitur ini menjadi representasi numerik dari karakteristik visual biji kopi dan dapat digunakan sebagai *input* pada algoritma klasifikasi seperti K-NN. Menurut (Han et al., 2012), ekstraksi fitur yang efektif memberikan informasi yang cukup akurat untuk membedakan citra, terutama jika digabungkan dengan klasifikasi berbasis *machine learning*. Dengan demikian, pengolahan citra digital, mulai dari segmentasi objek hingga ekstraksi fitur, merupakan metode yang relevan dan efektif dalam pendekatan klasifikasi biji kopi.

K-Nearest Neighbor (K-NN) adalah algoritma klasifikasi *non-parametrik* yang bekerja berdasarkan kedekatan data. K-NN menentukan kelas data uji berdasarkan suara mayoritas dari **k** tetangga terdekatnya dalam ruang fitur (Budianita et al., 2024). Metode ini memiliki keunggulan dalam kesederhanaannya dan toleransi yang relatif baik terhadap noise jika nilai k diatur dengan tepat. K-NN sangat bergantung pada metrik jarak (umumnya *Euclidean Distance*) untuk mengukur tingkat kesamaan antar fitur.

2.2.3 Naïve Bayes

Naive Bayes merupakan algoritma klasifikasi berbasis probabilistik yang bekerja berdasarkan **Teorema Bayes**. Metode ini mengasumsikan bahwa semua fitur bersifat independen terhadap satu sama lain, sebuah asumsi yang membuatnya sangat sederhana dan cepat dalam melakukan proses klasifikasi (Pemungkas dan Asnawi, 2024). Dalam konteks penelitian ini, Naive Bayes digunakan untuk mengklasifikasikan jenis biji kopi (Arabika atau Robusta) berdasarkan probabilitas fitur warna, tekstur, dan bentuk yang diperoleh dari citra.

Keunggulan utama Naive Bayes terletak pada efisiensi komputasinya, kecepatan dalam pelatihan model, dan performa yang tetap kompetitif meskipun asumsi independensi fitur sering kali tidak terpenuhi secara sempurna di dunia nyata (Sahar, 2020). Dalam studi oleh (Fata dan Avianto, 2024), Naive Bayes terbukti efektif menghasilkan prediksi yang layak untuk klasifikasi mutu biji kopi Robusta. Dalam penelitian ini, Naive Bayes akan mengklasifikasikan data berdasarkan probabilitas fitur yang diekstrak dari citra, dan hasil klasifikasi ini akan dibandingkan secara ketat dengan hasil dari algoritma K-NN.