

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kemajuan teknologi dan ilmu pengetahuan yang sangat pesat telah membawa dampak signifikan dalam berbagai sektor kehidupan, termasuk di bidang pertanian yaitu kelapa sawit. Kelapa sawit merupakan salah satu komoditas strategis di Indonesia yang berkontribusi besar terhadap perekonomian nasional (Fatima, 2024) . Keberhasilan industri kelapa sawit sangat ditentukan oleh kualitas bibit yang digunakan, karena bibit unggul menjadi fondasi utama dalam mencapai produktivitas tinggi, kualitas panen yang optimal, serta ketahanan terhadap hama dan penyakit. Sebaliknya, penggunaan bibit yang tidak sesuai dapat menyebabkan penurunan hasil panen, menurunnya mutu tandan buah segar, dan meningkatnya kerentanan terhadap serangan organisme pengganggu tanaman, yang pada akhirnya mengancam keberlanjutan dan daya saing industri ini

Indonesia dikenal sebagai negara dengan luas areal perkebunan kelapa sawit terbesar di dunia, yang tersebar di 22 dari 38 provinsi. Dua wilayah utama yang menjadi pusat pengembangan komoditas ini adalah Pulau Sumatera dan Kalimantan, di mana sekitar 90% total perkebunan kelapa sawit nasional berada di kedua pulau tersebut (Thakur et al., 2024). Salah satu wilayah potensial di Pulau Sumatera adalah Kabupaten Serdang Bedagai, Sumatera Utara, yang dikenal sebagai sentra produksi kelapa sawit dengan kontribusi signifikan terhadap pasokan minyak sawit nasional. Di wilayah ini, salah satu perusahaan perkebunan yang berperan aktif adalah PTPN IV Pabatu, yang menjalankan sistem pembibitan kelapa sawit melalui dua tahapan, yaitu tahap pembibitan awal dari umur 1- 3 bulan (*pre nursery*) dan pembibitan utama dari umur 4-12 bulan (*main nursery*). Proses pengelolaan pembibitan tersebut memegang peranan penting dan tidak dapat dianggap remeh, karena keberhasilan tanaman di lapangan sangat ditentukan oleh kualitas bibit yang dipersiapkan sejak dini.

Salah satu tantangan utama dalam proses seleksi bibit kelapa sawit pada tahap *pre nursery* dan *main nursery* adalah tingginya variasi genetik antar individu bibit serta perbedaan kemampuan adaptasi terhadap lingkungan. Kondisi ini menyebabkan performa pertumbuhan bibit menjadi sangat beragam. Sebagai contoh, bibit dengan karakteristik pertumbuhan yang unggul biasanya menunjukkan kualitas dan kuantitas yang mendukung peningkatan produktivitas tanaman kelapa sawit. Sebaliknya, bibit dengan pertumbuhan yang kurang unggul memerlukan perlakuan khusus, seperti penerapan strategi budidaya yang lebih intensif atau program perbaikan kualitas untuk dapat mencapai hasil yang optimal. Dalam konteks industri, bibit dengan kualitas rendah juga cenderung memiliki nilai jual yang lebih rendah, sehingga penting dilakukan klasifikasi sejak dini untuk menentukan penanganan yang tepat bagi setiap jenis bibit (Mahendra, 2023).

Untuk menjawab tantangan tersebut, dibutuhkan pendekatan teknologi yang mampu meningkatkan akurasi dan efisiensi dalam memprediksi kualitas bibit. Salah satu solusi yang dilakukan adalah pemanfaatan teknik *data mining*. *Data mining* adalah proses ekstraksi informasi penting dan berharga dari basis data untuk menghasilkan pengetahuan yang bermanfaat (Nurdin & Astika, 2019). Studi ini menggunakan algoritma klasifikasi *K-Nearest Neighbour* (KNN) dan *Naïve Bayes*, yang mudah diterapkan sehingga dapat memproses data dalam jumlah besar secara efisien dan mencapai akurasi tinggi bahkan pada data yang kompleks atau tidak lengkap (Putry, 2022).

Penerapan metode klasifikasi dalam bidang pertanian, khususnya dalam konteks seleksi bibit kelapa sawit di tahap pembibitan awal dan utama, algoritma KNN bekerja dengan memprediksi kualitas bibit dari data pengamatan pertumbuhan bibit dari umur 1-12 bulan berdasarkan kedekatan nilai kriteria seperti tinggi tanaman, diameter batang, panjang daun dan terserang hama pada tiap varietas bibit. Sementara itu, *Naïve Bayes* mengandalkan pendekatan probabilistik untuk menghitung kemungkinan suatu varietas bibit termasuk dalam suatu kategori berdasarkan distribusi historis dari masing-masing kriteria.

Penelitian ini memiliki urgensi strategis untuk mendukung proses seleksi bibit secara lebih objektif dan efisien, menggantikan metode konvensional yang cenderung subjektif. Di tengah persaingan global yang semakin ketat, kesalahan dalam pemilihan bibit dapat berdampak signifikan bahkan dapat menurunkan produktivitas hingga 30% (Damayanti et al., 2025). Oleh karena itu, penerapan algoritma KNN dan *Naïve Bayes* dalam klasifikasi kualitas varietas bibit kelapa sawit diharapkan dapat menjadi solusi inovatif dalam memprediksi kualitas varietas bibit unggul dan tidak unggul secara cepat dan akurat. Penelitian ini akan fokus pada kriteria utama: adalah umur bibit, tinggi bibit, diameter batang, jumlah daun, panjang daun dan terserang hama berdasarkan umur bibit, sebagai dasar klasifikasi dalam sistem berbasis *data mining* untuk mendukung peningkatan kualitas bibit kelapa sawit secara berkelanjutan.

Implementasi *data mining* pada berbagai studi kasus dan metodologi telah banyak dilakukan dalam penelitian terdahulu. Salah satu studi yang relevan adalah penelitian oleh (Nur et al., 2021) mengenai perbandingan metode K-NN dan *Naïve Bayes* untuk klasifikasi penentuan calon pendonor darah. Temuan dalam penelitian tersebut menunjukkan bahwa algoritma *K-Nearest Neighbor* memiliki performa yang lebih unggul dibandingkan dengan *Naïve Bayes*, di mana hasil pengujian mencatat tingkat akurasi sebesar 86% untuk K-NN, sementara *Naïve Bayes* hanya mencapai akurasi sebesar 76%.

Penelitian terdahulu yang dilakukan oleh (Nurdin, 2024) berfokus pada analisis *data mining* untuk memprediksi klasifikasi masyarakat kurang mampu dengan menerapkan metode *K-Nearest Neighbor* (K-NN). Penelitian ini bertujuan untuk menyediakan basis data prediktif mengenai keluarga prasejahtera yang berhak menerima program bantuan pemerintah. Prinsip kerja algoritma K-NN dalam studi ini dilakukan dengan mengukur jarak terdekat antara data uji terhadap data latih yang tersedia. Berdasarkan evaluasi menggunakan *Confusion Matrix*, model penelitian yang menggunakan rasio data latih dan data uji sebesar 70:30 (216 data latih dan 93 data uji) ini berhasil mencapai tingkat akurasi sebesar 86,02%. Selain itu, performa model juga diukur melalui parameter lain yang

menghasilkan nilai *Recall* sebesar 61,90%, Presisi 72,22%, dan *F1-Score* sebesar 66,04%.

Penelitian terkait juga dilakukan oleh (Puspitasari et al., 2022) dengan judul *Klasifikasi Kualitas Minyak Sawit Menggunakan Metode Klasifikasi Naïve Bayes*. Hasil penelitian menunjukkan bahwa metode klasifikasi *Naïve Bayes* mencapai tingkat akurasi yang cukup baik, yaitu 64,25% dengan tingkat kesalahan 35,7%. Hasil ini menunjukkan bahwa metode klasifikasi *Naïve Bayes* memberikan hasil yang baik dalam mengklasifikasikan kualitas minyak kelapa sawit.

Penelitian lain yang telah dilakukan (Valentino Jayadi et al., 2023) dengan judul “Perbandingan *K-Nearest Neighbor* dan *Support Vector Machine* untuk Klasifikasi Kualitas Udara di Jakarta”. Studi ini bertujuan untuk membandingkan akurasi algoritma *K-Nearest Neighbor* (KNN) dan *Support Vector Machine* (SVM) dalam mengklasifikasikan kualitas udara di wilayah Jakarta. Hasil penelitian menunjukkan bahwa algoritma *Support Vector Machine* (SVM) dengan kernel fungsi basis radial (RBF) dan parameter kernel 100 memiliki kinerja yang lebih baik daripada algoritma KNN dengan nilai $K = 6$. Metode SVM mencapai akurasi 98%, presisi 97%, tingkat pencarian kembali 97%, dan skor F1 97%, sedangkan metode KNN mencapai akurasi 96%, presisi 96%, tingkat penemuan kembali 93%, dan skor F1 94%.

Penelitian terdahulu yang dilakukan oleh (Tangkelayuk, 2022) berfokus pada klasifikasi kualitas air dengan melakukan studi komparatif antara tiga algoritma, yaitu *K-Nearest Neighbor* (K-NN), *Naïve Bayes*, dan *Decision Tree*. Tujuan utama dari penelitian tersebut adalah untuk mengevaluasi performa masing-masing metode dalam menentukan tingkat akurasi tertinggi pada klasifikasi data. Hasil pengujian menunjukkan bahwa algoritma K-NN memberikan performa terbaik dengan tingkat akurasi sebesar 86,88%, yang secara signifikan mengungguli metode *Decision Tree* dengan akurasi 80,84% dan *Naïve Bayes* sebesar 63,60%. Berdasarkan hasil capaian tersebut, penelitian tersebut menyimpulkan bahwa K-NN merupakan metode yang paling efektif untuk diterapkan dalam klasifikasi data kualitas air dibandingkan kedua algoritma lainnya.

Berdasarkan penjelasan dari penelitian sebelumnya, dan menjadi referensi pendukung, maka penulis mengambil judul “Perbandingan Metode *K-Nearest Neighbor* dan Metode *Naïve Bayes* pada Klasifikasi Kualitas Varietas Bibit Kelapa Sawit”. Penelitian ini berfokus memprediksi kualitas bibit dengan melakukan klasifikasi terhadap varietas bibit kelapa sawit, dengan objek penelitian berupa data pengamatan pertumbuhan dari bibit kelapa sawit dari umur 1-12 bulan pada tiap varietas. Berdasarkan analisis beberapa penelitian sebelumnya yang relevan, belum ditemukan penelitian serupa yang membahas topik ini secara spesifik. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi penting bagi industri kelapa sawit, terutama dalam meningkatkan kualitas dan daya saing pada jenis bibit yang dihasilkan.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan masalah yang akan dibahas dalam penelitian ini adalah :

1. Bagaimana penerapan metode *K-Nearest Neighbor* dan *Naïve Bayes* dalam proses klasifikasi kualitas varietas bibit kelapa sawit menjadi kelas unggul dan tidak unggul pada PTPN IV Pabatu.
2. Bagaimana perbandingan kinerja metode *K-Nearest Neighbor* dan *Naïve Bayes* berdasarkan nilai metrik evaluasi akurasi, presisi, *Recall*, dan *F1-Score* dalam mengklasifikasikan kualitas varietas bibit kelapa sawit.

1.3 Tujuan Penelitian

Berikut merupakan tujuan dari penelitian berdasarkan uraian masalah diatas:

1. Menerapkan metode *K-Nearest Neighbor* dan *Naïve Bayes* dalam mengklasifikasikan kualitas varietas bibit kelapa sawit ke dalam kelas unggul dan tidak unggul berdasarkan data pengamatan pertumbuhan bibit pada PTPN IV Pabatu.
2. Membandingkan kinerja metode *K-Nearest Neighbor* dan *Naïve Bayes* berdasarkan nilai metrik evaluasi akurasi, presisi, *Recall*, dan *F1-Score*

untuk menentukan metode yang memiliki performa terbaik dalam mengklasifikasikan kualitas varietas bibit kelapa sawit.

1.4 Manfaat Penelitian

Dalam penelitian ini, penulis diharapkan mampu memberikan beberapa manfaat, antara lain sebagai berikut:

1. Memberikan kontribusi dalam pengembangan penelitian data mining melalui penerapan algoritma *K-Nearest Neighbor* dan *Naïve Bayes*, khususnya dalam klasifikasi data pada bidang pertanian.
2. Meningkatkan pemahaman mengenai penerapan algoritma *K-Nearest Neighbor* dan *Naïve Bayes* dalam klasifikasi kualitas varietas bibit kelapa sawit guna mendukung kelayakan bibit, serta menjadi acuan bagi penelitian selanjutnya di sektor perkebunan.

1.5 Ruang Lingkup dan Batasan Penelitian

Untuk lebih memfokuskan penelitian, penulis membatasi masalah yang digunakan dalam penelitian ini. Adapun batasan masalah pada penelitian ini adalah sebagai berikut :

1. Penelitian ini secara khusus mengimplementasikan algoritma *K-Nearest Neighbor* (K-NN) dan *Naïve Bayes* sebagai instrumen utama dalam melakukan klasifikasi data.
2. Objek penelitian ini terdiri atas 500 pokok bibit kelapa sawit, yang mencakup 10 varietas berbeda dengan masing-masing varietas berjumlah 50 pokok. Data penelitian diperoleh dari hasil pengamatan pertumbuhan bibit kelapa sawit yang dilaksanakan pada bulan Maret 2025 di PTPN IV Pabatu.
3. Kriteria atau fitur yang digunakan meliputi variabel umur bibit, tinggi bibit, diameter batang, jumlah daun, panjang daun, serta status serangan hama yang dikategorikan berdasarkan varietas bibit masing-masing.

4. Dataset yang berjumlah 500 dilakukan skema pembagian data (*data splitting*) dengan rasio 70% untuk data latih (*training data*) dan 30% untuk data uji (*testing data*).
5. Proses klasifikasi dalam penelitian ini dibatasi hanya pada dua kelas, yaitu kelas unggul dan tidak unggul. Kelas tidak unggul merepresentasikan bibit yang belum memenuhi standar kualitas yang ditetapkan, bukan bibit yang mengalami kerusakan fisik atau kematian.
6. Perancangan dan implementasi sistem dibangun dengan menggunakan bahasa pemrograman *Python* dan *database* SQLite sebagai media penyimpanan data.