

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Indonesia merupakan negara dengan dua musim yang memiliki kekayaan berlimpah di bidang perkebunan dan pertanian sebagai penghasil produk unggulan hortikultura yang tinggi [1]. Salah satu produk pada bidang tersebut adalah buah-buahan dan sayur-sayuran, termasuk tomat. Tanaman ini dapat tumbuh dengan subur di berbagai tempat dan merupakan salah satu alasan pilihan petani dalam membudidayakannya. Namun, terkadang hasil panen dari buah yang satu ini masih belum memuaskan seperti yang diharapkan. Beberapa penyebab gagalnya hasil panen tersebut diantaranya adalah teknik budidaya, kondisi lingkungan dan gangguan hama serta penyakit. Salah satu penyakit utama pada tanaman tomat adalah bercak daun [2].

Penyakit tanaman dapat menyebabkan penurunan produksi pertanian, sehingga pendeteksian dini dan diagnosis penyakit tanaman menjadi sangat penting. Umumnya, penyakit pada tomat sering muncul pada daun dengan ciri-ciri daun yang terserang bisa bermacam-macam dan sulit dibedakan. Selain itu, penyakit tanaman tersebut tidak hanya berbahaya bagi manusia, tetapi juga bagi hewan [3]. Oleh karena itu, deteksi dan analisis penyakit pada tanaman tomat untuk meningkatkan hasil panen sangat penting dilakukan, karena sangat sulit untuk mendeteksi dan menganalisis penyakit tomat secara manual [4]. Para peneliti telah banyak melakukan penelitian untuk memperoleh keefektifan dalam mengidentifikasi penyakit tanaman [5]. Aplikasi berbasis pengolahan citra untuk pengenalan dan klasifikasi penyakit tanaman merupakan bidang penelitian yang luas saat ini. Aplikasi ini berguna untuk identifikasi atau pengenalan penyakit tanaman secara tepat, seperti: jamur, bakteri dan virus yang dapat merusak tanaman [6].

Beberapa penelitian telah menghasilkan metode yang tepat di dalam melakukan pengenalan atau identifikasi penyakit tomat dengan berbagai teknik, salah satunya menggunakan teknik *convolutional neural network* (CNN) dengan hasil akurasi yang baik di atas 90% [7]. Teknik lainnya menggunakan kombinasi *random forest* and *k-nearest neighbor* (KNN) dengan akurasi mencapai 96% [8]. Selain itu, klasifikasi penyakit pada tomat dapat dilakukan dengan teknik *classification tree* dengan akurasi ketepatan pengenalan mencapai 97.3% [6]. Pada dasarnya, setiap teknik identifikasi atau pengenalan

yang dilakukan harus disertai dengan data citra, teknik pengolahan citra dan ekstraksi ciri yang tepat untuk memperoleh hasil yang terbaik.

Proses pengenalan citra di dalam ilmu komputer dikenal dengan istilah *computer vision* yang dapat diterapkan pada bidang pertanian dalam melakukan identifikasi penyakit tanaman [9]. Teknik *machine learning* yang umum digunakan dalam identifikasi citra [10], harus dimulai dengan memperoleh data citra yang tidak mengandung banyak *noise* (*background*) untuk menjaga hasil identifikasi yang baik [11]. Dalam hal ini, segmentasi citra menjadi bagian penting pada teknik pengolahan citra dalam beberapa penelitian untuk identifikasi citra [12]. Segmentasi citra merupakan proses pemisahan antara objek penting dengan *background* dari suatu citra [13] dan merupakan dasar dari *computer vision*, dimana akurasi segmentasi memiliki dampak yang besar terhadap hasil identifikasi yang dilakukan [14]. Keluaran dari proses segmentasi citra adalah citra yang tersegmentasi untuk beberapa kelas, dimana setiap kelas memiliki atributnya masing-masing [15]. Berbagai teknik segmentasi dapat digunakan dan umumnya menggunakan teknik *clustering* dalam memisahkan antara objek dengan bagian *background* suatu citra. Salah satu teknik yang paling populer di dalam segmentasi citra adalah *k-means clustering* yang telah banyak diusulkan dalam meningkatkan akurasi pengenalan objek berbasis citra.

Algoritma pengelompokan *k-means clustering* dimulai dengan memilih nilai k (jumlah *cluster*) dan nilai *centroid* awal secara acak untuk setiap *cluster*. Pada setiap iterasi, objek data diasosiasikan dengan *centroid* terdekat serta diperbarui berdasarkan nilai jarak objek data yang terdekat [16]. Algoritma *k-means clustering* dalam segmentasi citra merupakan algoritma *unsupervised learning* yang digunakan untuk mensegmentasi suatu objek citra dengan bagian *background* [17], sehingga diperlukan pengaturan jumlah pengelompokan segmentasi dan *centroid* awal yang memberikan pengaruh baik atau buruk pada kualitas segmentasi [18]. Permasalahan umum yang sering dijumpai adalah penentuan *centroid* awal yang dilakukan secara acak, sehingga menghasilkan *cluster* yang berbeda-beda jika inisialisasi awalnya diubah [11]. Perubahan tersebut dapat menyebabkan gagalnya citra tersegmentasi dengan baik dan mempengaruhi hasil identifikasi yang dilakukan [19]. Solusi permasalahan ini dapat dilakukan dengan menggunakan algoritma *k-means clustering* yang ditingkatkan untuk pemisahan antara citra yang diteliti dan *background* secara tepat [20].

Berdasarkan penjelasan sebelumnya, proses segmentasi pada identifikasi penyakit tomat sangat diperlukan agar dapat meningkatkan hasil akurasi. Pada penelitian ini, akan

dilakukan identifikasi citra daun tomat yang terinfeksi penyakit menggunakan algoritma KNN (*K-Nearest Neighbor*) berdasarkan segmentasi menggunakan algoritma peningkatan pada *k-means clustering* untuk memisahkan bagian objek dengan *background*. Kontribusi pada penelitian ini terletak pada proses segmentasi citra yang dilakukan menggunakan metode *cosine similarity* pada algoritma *k-means clustering* dalam penentuan *centroid* awal. *Cosine similarity* merupakan metode kemiripan pada data yang menghasilkan tingkat kemiripan dengan nilai error yang lebih kecil dibandingkan metode *Euclidean* dan *Jaccard Similarity* [21].

Cosine similarity akan mencari nilai fitur dengan kemiripan tertinggi dan terendah pada setiap data citra berdasarkan kandungan warna citra. Nilai ini akan digunakan sebagai *centroid* awal pada algoritma *k-means clustering* dalam melakukan segmentasi citra. Pemisahan fitur ini dilakukan agar data dapat dikelompokkan sesuai nilai fitur yang sama dengannya. Setiap algoritma pengelompokan menginduksi kesamaan antara titik data yang diberikan, sesuai dengan kriteria pengelompokan yang mendasarinya [22]. Pada dasarnya teknik pengelompokan data dilakukan dengan menguji kemiripan antar satu fitur dengan fitur pusatnya. Semakin dekat ukuran kemiripan suatu fitur dengan fitur pusatnya, maka fitur tersebut akan dikelompokkan pada *cluster* yang sama dan sebaliknya demikian.

Pada penelitian ini, penggunaan *cosine* akan memperoleh beberapa karakteristik fitur dengan kemiripan yang berbeda antara satu dengan lainnya. Nilai fitur tersebut menjadi dasar di dalam proses segmentasi yang diusulkan dan selanjutnya algoritma *k-means clustering* akan mencari kesesuaian fitur tersebut dan mempartisinya menjadi beberapa bagian atau segmentasi citra. Semakin baik nilai *centroid* awal yang dipilih pada algoritma *k-means clustering*, maka semakin baik pula hasil *clustering* yang dihasilkan [23]. Hasil akhir *clustering* yang baik akan menemukan segmentasi citra yang sesuai dan meningkatkan hasil identifikasi yang dilakukan.

Berdasarkan pertimbangan tersebut di atas, maka tugas akhir ini akan membahas mengenai penerapan metode KNN untuk mengidentifikasi penyakit tanaman tomat dan perbandingan beberapa metode ekstraksi fitur, sehingga tugas akhir ini diberi judul **“Identifikasi Penyakit Tomat Menggunakan Metode KNN Berdasarkan Segmentasi Enhanced K-Means Clustering”**

1.2 Rumusan Masalah

Penelitian ini akan membahas tentang identifikasi penyakit tanaman tomat yang berdasarkan citra dengan menggunakan Algoritma KNN dan Segmentasi *Enhanced K-means Clustering* dan beberapa metode ekstraksi fitur.

1.3 Pembatasan Masalah

Batasan masalah pada pembahasan Tugas Akhir ini yaitu sebagai berikut:

1. Data set berupa daun dari setiap penyakit tanaman tomat diambil dari Kaggle (<https://www.kaggle.com/kaustubhb999/tomatoleaf>) yang terdiri dari 5 jenis penyakit yaitu *Bacterial Spot*, *Spider Mites*, *Target Spot*, *Mosaic Virus*, *Yellow L.C Virus*.
2. Ukuran citra yang digunakan berukuran 256 x 256 dan menggunakan format *.JPG.
3. Ekstraksi ciri yang digunakan pada penelitian ini adalah metode GLCM dan HSV

1.4 Tujuan dan Manfaat

1.4.1 Tujuan

Tujuan dari penelitian ini adalah untuk mendeteksi penyakit tanaman tomat dengan menggunakan algoritma KNN dan segmentasi *Enhanced K-Means Clustering* serta membandingkan beberapa metode ekstraksi fitur dan warna sehingga dapat mengetahui metode ekstraksi yang paling bagus dalam mendeteksi penyakit daun pada tanaman tomat.

1.4.2 Manfaat

Adapun manfaat dari penelitian ini adalah:

1. Membantu petani dalam mendeteksi penyakit pada tanaman tomat dengan cepat tanpa bantuan tenaga ahli.
2. Mengetahui cara penerapan algoritma KNN dan segmentasi *Enhanced K-Means Clustering* dalam mendeteksi penyakit.

1.5 Keterbaruan

Pada penelitian ini, akan dilakukan identifikasi citra daun tomat yang terinfeksi penyakit menggunakan algoritma KNN (*K-Nearest Neighbor*) berdasarkan segmentasi menggunakan algoritma peningkatan pada *k-means clustering* untuk memisahkan bagian objek dengan *background*. Kontribusi pada penelitian ini terletak pada proses segmentasi citra yang dilakukan menggunakan metode *cosine similarity* pada algoritma *k-means clustering* dalam penentuan *centroid* awal.

Cosine similarity akan mencari nilai fitur dengan kemiripan tertinggi dan terendah pada setiap data citra berdasarkan kandungan warna citra. Nilai ini akan digunakan sebagai *centroid* awal pada algoritma *k-means clustering* dalam melakukan segmentasi citra. Pemisahan fitur ini dilakukan agar data dapat dikelompokkan sesuai nilai fitur yang sama dengannya. Setiap algoritma pengelompokan menginduksi kesamaan antara titik data yang diberikan, sesuai dengan kriteria pengelompokan yang mendasarinya. Pada dasarnya teknik pengelompokan data dilakukan dengan menguji kemiripan antar satu fitur dengan fitur pusatnya. Semakin dekat ukuran kemiripan suatu fitur dengan fitur pusatnya, maka fitur tersebut akan dikelompokkan pada *cluster* yang sama dan sebaliknya demikian.

Penggunaan *cosine* akan memperoleh beberapa karakteristik fitur dengan kemiripan yang berbeda antara satu dengan lainnya. Nilai fitur tersebut menjadi dasar di dalam proses segmentasi yang diusulkan dan selanjutnya algoritma *k-means clustering* akan mencari kesesuaian fitur tersebut dan mempartisinya menjadi beberapa bagian atau segmentasi citra. Semakin baik nilai *centroid* awal yang dipilih pada algoritma *k-means clustering*, maka semakin baik pula hasil *clustering* yang dihasilkan. Hasil akhir *clustering* yang baik akan menemukan segmentasi citra yang sesuai dan meningkatkan hasil identifikasi yang dilakukan.