

BAB I

PENDAHULUAN

1.1 Latar Belakang

Citra merupakan istilah lain dari gambar yang memberikan informasi secara visual. Identifikasi citra pada dataset yang berjumlah kecil masih dapat dilakukan secara manual. Namun, ketika data citra yang jumlahnya banyak, proses identifikasi diperlukan bantuan komputer untuk mengurangi kesalahan yang dilakukan manusia. Terkadang dalam sebuah dataset, tidak semua data yang masuk merupakan data yang normal. Akan tetapi, ada beberapa data yang tidak termasuk dalam kategori pada dataset tersebut. Biasanya jumlah dari data tersebut pada umumnya berjumlah sedikit sekali. Data tersebut dikenal dengan nama *outlier* atau anomali.

Outlier merupakan data yang dianggap memiliki karakteristik yang berbeda dibandingkan dengan kebanyakan data lainnya [1]. Hal ini terjadi dikarenakan adanya kesalahan pengukuran, kesalahan pada pengisian data, sumber data yang berbeda dan peristiwa langka [2]. Data *outlier* yang berada di dataset akan mempengaruhi terhadap hasil dari proses selanjutnya seperti *clustering* dan klasifikasi dan terkadang *outlier* juga menyimpan informasi yang penting sehingga dapat digunakan untuk mendeteksi perilaku atau kejadian yang tidak normal pada suatu masalah seperti deteksi kecurangan kartu kredit, deteksi penyusupan jaringan, monitoring aktivitas, gangguan ekosistem alam, bidang kesehatan, deteksi data yang tidak sesuai dengan data pelatihan dan lain-lain [1][2]. Oleh karena itu, dibutuhkan algoritma yang dapat melakukan pendeteksian *outlier* agar dataset bersih dari *outlier*.

Beberapa pendekatan untuk mendeteksi *outlier* antara lain *Statistical Approaches*, *Proximity-based Approaches*, *Cluster-based Approaches*, *Distance Approaches* dan *Classification Approaches* [1]. Tidak semua pendekatan tersebut mampu untuk menemukan *outlier* dengan tingkat akurasi yang tinggi. Dengan hal itu maka pendekatan berbasis *cluster* sangat cocok untuk menyelesaikan permasalahan tersebut [3].

Deteksi *outlier* dengan pendekatan berbasis *cluster* merupakan teknik dari *clustering* yang berusaha untuk menemukan objek-objek yang berelasi secara kuat, sedangkan untuk deteksi

outlier berusaha untuk menemukan objek-objek yang berelasi dengan objek lainnya secara lemah yaitu berada di *cluster* kecil dan berada jauh dari pusat *cluster* [1]. Adapun pendekatan berbasis *cluster* antara lain *Cluster Based Local Outlier Factor* (CBLOF), *Local Density Cluster-Based Outlier Factor* (LDCOF) dan *unweighted Cluster Based Local Outlier Factor* (*unweighted-CBLOF*).

Algoritma *unweighted-CBLOF* merupakan Algoritma yang terbaik dari pendekatan berbasis *cluster* berdasarkan tingkat akurasi dalam menemukan *outlier*, kecepatan proses terhadap dataset yang besar dan juga mampu mendeteksi *outlier* yang sifatnya global [3]. Penerapan algoritma *unweighted-CBLOF* dalam menentukan data normal dan *outlier* sangatlah bergantung dari hasil *clustering*[4]. Algoritma *K-Means* merupakan algoritma yang telah digunakan untuk proses *clustering* sebelum deteksi *outlier* dengan *unweighted-CBLOF* [5]. Namun, algoritma *K-Means* mempunyai kelemahan jika digunakan pada dataset yang terdapat *outlier*[6]. Dengan permasalahan tersebut, maka dibutuhkan suatu algoritma yang mampu memberikan hasil *cluster* yang baik.

Clustering dengan pendekatan berbasis medoid merupakan salah satu solusi untuk menyelesaikan permasalahan mengenai keberadaan *outlier* pada dataset[6]. Adapun algoritma *clustering* yang berbasis medoid antara lain PAM, CLARA dan CLARANS. Dari ketiga algoritma tersebut CLARANS dapat dijadikan pengganti dari algoritma *K-Means*. Hal ini dikarenakan CLARANS lebih cepat dibandingkan algoritma PAM dalam menangani data yang berukuran besar [7] dan efisien dalam menemukan *outlier* ketika digunakan untuk proses *clustering* sebelum deteksi *outlier* [8].

Berdasarkan permasalahan diatas, penelitian ini akan menggunakan ekstraksi fitur warna berbasis statistik orde pertama dan tekstur dengan *Gray Level Co-occurrence Matrix* (GLCM). Sedangkan untuk proses *clustering* menggunakan algoritma CLARANS. Kemudian hasil *clustering* akan di deteksi *outlier* dengan menggunakan *unweighted-CBLOF*. Penelitian ini diharapkan mendapatkan hasil yang maksimal dalam menemukan *outlier* di dataset.

1.2 Perumusan Masalah

Berdasarkan latar belakang yang ada diatas, maka rumusan masalah yang akan uji dalam penelitian ini adalah :

1. Fitur apa yang terbaik diantara fitur tekstur, fitur warna dan fitur warna dengan tekstur untuk deteksi *outlier* ?
2. Bagaimana pengaruh penentuan jumlah kelas pada Algoritma CLARANS terhadap hasil akurasi deteksi *outlier* ?

1.3 Tujuan

Tujuan dari penelitian ini adalah :

1. Mengetahui fitur yang terbaik diantara fitur tekstur, fitur warna dan fitur warna dengan tekstur untuk deteksi *outlier*.
2. Mengetahui pengaruh penentuan jumlah kelas pada Algoritma CLARANS terhadap hasil akurasi deteksi *outlier*.

1.4 Manfaat

Hasil dari penelitian ini nantinya dapat dimanfaatkan untuk deteksi *outlier* pada kasus yang datanya berbentuk citra dan sebagai bahan referensi untuk penelitian selanjutnya terkait dengan deteksi *outlier* pada data citra.

1.5 Batasan Masalah

Berdasarkan uraian yang telah dijelaskan di atas maka dibuatlah batasan masalah agar penelitian fokus terhadap topik permasalahan. Adapun batasan masalah pada permasalahan ini sebagai berikut :

1. Dataset yang digunakan adalah Dataset *Supermarket Produce* [9] sebagai data normal dan 50 gambar dari dataset *Fruit Image* [10] sebagai data *outlier*.
2. Proses pemotongan daerah yang penting pada citra dilakukan secara manual.
3. Perangkat lunak yang digunakan adalah Spyder 3.2.3.
4. Skenario uji coba meliputi,
 - a. Penggunaan ekstraksi fitur tekstur, fitur warna dan fitur warna dengan tekstur.
 - b. Jumlah kelas yang digunakan 15,20 dan 25
 - c. Jumlah data *outlier* 10,20,30,40 dan 50 pada setiap skenario.

1.6 Metodologi Penelitian

Secara umum, penelitian dilakukan dengan beberapa tahapan, yaitu:

- a. Studi Literatur
Studi literatur yaitu melakukan pengumpulan informasi yang diperlukan untuk penelitian skripsi. Informasi yang diperlukan diperoleh dengan mempelajari literatur- literatur mengenai deteksi *outlier*, model ruang HSV, ekstraksi fitur warna dengan perhitungan statistika orde pertama dan ekstraksi fitur tekstur dengan *Gray Level Co-occurrence Matrix* (GLCM), algoritma CLARANS dan algoritma *unweighted-CBLOF*.
- b. Pengumpulan Data dan *Pre prosesing* Data
Pengumpulan dataset yaitu melakukan pencarian data yang akan digunakan untuk penelitian seperti yang dipaparkan di batasan masalah. *Pre prosesing* data yaitu pemotongan daerah yang penting pada data citra.
- c. Perancangan Sistem
Perancangan sistem yaitu melakukan pembuatan desain sistem *Outlier* berdasarkan algoritma yang telah ditentukan. yang diimplementasikan. Alur dari desain tersebut meliputi proses ekstraksi fitur warna berbasis statistik orde pertama dan ekstraksi fitur tekstur dengan *Gray Level Co-occurrence Matrix* (GLCM). Proses selanjutnya, *clustering* menggunakan algoritma *Clustering Large Applications Based on Randomized Search* (CLARANS) pada data hasil ekstraksi fitur. Setelah data sudah dicluster kemudian dilakukan deteksi *Outlier* menggunakan algoritma *unweighted-CBLOF*. Hasilnya dari sistem ini berupa data yang diidentifikasi sebagai *Outlier* dan normal.
- d. Implementasi Sistem Deteksi *Outlier*
Proses implementasi sistem deteksi *outlier* merupakan implementasi dari rancangan sistem ke dalam suatu kode program. Bahasa pemrograman yang digunakan adalah Python.
- e. Uji Coba Sistem
Proses uji coba sistem berdasarkan skenario yang telah tentukan.
- f. Analisa
Analisa sistem untuk mengetahui tingkat akurasi dari proses deteksi *outlier* dengan menggunakan *unweighted-CBLOF* dengan *Clustering Large Applications Based on Randomized*

- Search* (CLARANS) dalam mengidentifikasi *outlier* berdasarkan fitur dan jumlah kelas yang digunakan.
- g. Penyusunan Laporan
Pada tahap ini dilakukan penyusunan laporan akhir dari semua tahap yang telah dilakukan.

1.7 Sistematika Penulisan

Adapun sistematika pembahasan skripsi ini adalah sebagai berikut:

BAB I PENDAHULUAN

Pendahuluan berisikan latar belakang, rumusan masalah, tujuan, manfaat penelitian, metodologi penelitian dan sistematika penulisan. Latar belakang mengenai topik deteksi *Outlier* serta algoritma yang akan digunakan yaitu CLARANS dan *unweighted-CBLOF*. Rumusan masalah dan tujuan penelitian yaitu untuk mengetahui pengaruh penggunaan fitur warna dan tekstur serta penentuan jumlah kelas terhadap hasil akurasi deteksi *outlier*. Batasan masalah berisikan batas-batasan untuk penelitian yang akan dilakukan.

BAB II STUDI PUSTAKA

Studi pustaka berisikan mengenai bahan atau materi yang diambil dan menjadi acuan dalam penelitian yang sumbernya dari berbagai penelitian terdahulu. Bahan dan materi yaitu mengenai citra digital, model ruang warna HSV, ekstraksi fitur dengan statistika orde pertama dan *Gray Level Co-occurrence Matrix* (GLCM), algoritma *clustering* CLARANS dan algoritma deteksi *outlier unweighted-CBLOF*.

BAB III PERANCANGAN SISTEM

Perancangan sistem berisikan tentang alur rancangan sistem deteksi *outlier* secara keseluruhan, mulai dari *pre processing*, ekstraksi fitur, *clustering* dan deteksi *outlier*.

BAB IV IMPLEMENTASI SISTEM

Implementasi menjelaskan mengenai implementasi sistem deteksi *outlier* ke dalam bahasa pemrograman dan diuji coba berdasarkan skenario yang telah ditentukan. Skenario uji coba meliputi penggunaan fitur dan penentuan jumlah kelas.

BAB V PENUTUP

Pada bab ini berisikan kesimpulan dan saran secara keseluruhan mengenai penelitian deteksi *outlier* yang telah dilakukan.