

BAB II

LANDASAN TEORI

2.1 State Of The Art

Penelitian terdahulu ini menjadi salah satu acuan dalam melakukan penelitian sehingga dapat memperkaya teori yang digunakan dalam mengkaji penelitian yang dilakukan. Ulasan penelitian terkait, dilakukan dengan maksud untuk menganalisis penelitian yang telah dilakukan sebelumnya. Penelitian terdahulu dapat dilihat sebagai judul berikut :

Dalam jurnal yang berjudul Klasifikasi Jenis Ikan Koi Menggunakan Gray Level Cooccurrence Matrix Dan Algoritma Naive Bayes. Tri Adhi Atmaji, Catur Supriyanto (2019), Penelitian ini bertujuan mengetahui jenis dari ikan koi tersebut. Masalah pengenalan jenis ikan koi dapat diselesaikan dengan klasifikasi menggunakan ekstraksi fitur GLCM (Gray Level Co-occurrence Matrix) dan algoritma Naive Bayes. Berdasarkan hasil penelitian yang telah dilakukan dengan 6 jenis ikan koi yaitu Asagi, Kohaku, Showa, Sanke, Shusui dan Tancho mendapatkan nilai akurasi sebesar 73,33% dengan menggunakan parameter jarak $d = 2$ pada sudut $\theta = 45^\circ$, dengan jumlah data benar 22 dan jumlah data salah 8.

Dalam jurnal yang berjudul Analisis Identifikasi Pola Ikan Koi Menggunakan Metode Sobel Edge Detection. Arif Wicaksono (2017), Penelitian ini bertujuan Mendeteksi Jenis ikan Koi dengan menggunakan Gambar/Citra dapat di jadikan solusi untuk mengetahui jenis ikan Koi. Gambar/Citra ikan Koi akan mengalami proses Greyscale untuk mendapatkan nilai keabuan kemudian akan mengalami

proses deteksi tepi dengan Sobel Edge Detection untuk mengetahui pola Ikan Koi. Setelah mendapatkan pola warna Ikan Koi aplikasi akan mencocokkan jenis ikan Koi berdasarkan nilai RGB dengan data jenis Ikan Koi yang ada di Database menggunakan K-Nearest Neighbor. Kelebihan dari cara pengenalan ikan Koi menggunakan Gambar/Citra ini adalah dapat mengenali jenis ikan koi secara cepat dan mengurangi penipuan yang dilakukan para penjual ikan Koi yang tidak bertanggung jawab.

Dalam jurnal yang berjudul Identifikasi Varietas Koi Berdasarkan Gambar Menggunakan Zero Parameter Simple Linear Iterative Clustering dan Support Vector Machine. Amadea Sapphira, dkk. (2012), Penelitian ini dilakukan pengujian pada dua metode superpixel segmentation, yaitu metode Simple Linear Iterative Clustering (SLIC) dan Zero Parameter Simple Linear Iterative Clustering (SLICO) untuk proses segmentasi objek untuk menghilangkan background pada gambar. Metode Color Local Binary Pattern digunakan untuk mendapatkan tekstur pada gambar melalui colorspace RGB, HSV, dan Grayscale. Support Vector Machine (SVM) digunakan untuk mengidentifikasi jenis ikan koi. Untuk menguji SVM, digunakan dua jenis kernel, yaitu kernel linear dan kernel Radial Basis Function (RBF). Hasil dari penelitian ini adalah program dapat mengenali jenis ikan koi pada gambar. Hasil pengujian menunjukkan akurasi sebesar 36% pada colorspace grayscale, 50% pada colorspace RGB, dan 48% pada colorspace HSV.

Dalam jurnal yang berjudul Analisis Identifikasi Pola Warna Ikan Koi Menggunakan Metode Sobel Edge Detection Dalam Karakteristik Citra Sharpening. Milfa Yetri, dkk. (2020), Peneliti ini bertujuan untuk membedakan

antara objek dengan latar belakang yang ada pada sebuah citra, sehingga proses ini dapat diimplementasikan untuk kepentingan tertentu, seperti pengenalan pola, deteksi wajah pada kamera dan sebagainya. Beberapa metode pendeteksian tepi memiliki kemampuan yang berbeda-beda dalam proses pengolahan citra untuk mendapatkan pola garis tepi. Salah satu metode yang digunakan pada penelitian ini adalah metode Sobel dengan objek penelitian menggunakan citra ikan koi. Kemampuan metode ini dibuktikan dengan menganalisis pola warna ikan yang terdapat pada objek citra yang telah mengalami proses karakteristik citra Shapening (penajaman citra). Hasilnya metode Sobel mendapatkan pola garis tepi warna ikan koi lebih baik dan akurat.

Dalam jurnal yang berjudul Identifikasi Jenis Ikan Menggunakan Model Hybrid Deep Learning Dan Algoritma Klasifikasi. Anifuddin Azis. (2020), Peneliti ini bertujuan untuk Untuk menentukan berapa jumlah spesies tersebut dibutuhkan suatu keahlian di bidang taksonomi. Dalam pelaksanaannya mengidentifikasi suatu jenis ikan bukanlah hal yang mudah karena memerlukan suatu metode dan peralatan tertentu, juga pustaka mengenai taksonomi. Pemrosesan video atau citra pada data ekosistem perairan yang dilakukan secara otomatis mulai dikembangkan. Dalam pengembangannya, proses deteksi dan identifikasi spesies ikan menjadi suatu tantangan dibandingkan dengan deteksi dan identifikasi pada objek yang lain. Metode deep learning yang berhasil dalam melakukan klasifikasi objek pada citra mampu untuk menganalisis data secara langsung tanpa adanya ekstraksi fitur pada data secara khusus. Sistem tersebut memiliki parameter atau bobot yang berfungsi sebagai ekstraksi fitur maupun sebagai pengklasifikasi. Data yang diproses

menghasilkan output yang diharapkan semirip mungkin dengan data output yang sesungguhnya. CNN merupakan arsitektur deep learning yang mampu mereduksi dimensi pada data tanpa menghilangkan ciri atau fitur pada data tersebut. Pada penelitian ini akan dikembangkan model hybrid CNN (Convolutional Neural Networks) untuk mengekstraksi fitur dan beberapa algoritma klasifikasi untuk mengidentifikasi spesies ikan. Algoritma klasifikasi yang digunakan pada penelitian ini adalah : Logistic Regression (LR), Support Vector Machine (SVM), Decision Tree, K-Nearest Neighbor (KNN), Random Forest, Backpropagation.

Dalam jurnal yang berjudul Analisis Matthew Correlation Coefficient pada K-Nearest Neighbor dalam Klasifikasi Ikan Hias. Novia Hasdyna, dkk. (2020), Penelitian ini menganalisis kinerja algoritma klasifikasi KNN dalam pengklasifikasian data ikan hias. Hasil penelitian ini menunjukkan bahwa nilai matthew correlation coefficient pada k-nn dengan menggunakan Euclidean distance diperoleh nilai MCC tertinggi pada class medium sebesar 0,786542. Nilai mcc tertinggi kedua pada class premium senilai 0,567434. Nilai mcc terendah adalah pada class low sebesar 0,435269. Dengan demikian, dapat dikatakan nilai korelasi matthew yang paling baik adalah pada class medium karena class medium adalah yang paling mendekati +1. Sebaliknya nilai MCC pada class low adalah yang paling mendekati 0. Adapun rata-rata nilai MCC pada klasifikasi data ikan hias dengan KNN adalah senilai 0,596415. Oleh karena itu, kinerja algoritma klasifikasi KNN adalah baik.

Dalam jurnal yang berjudul Klasifikasi Jenis Ikan Cupang Menggunakan Algoritma Principal Component Analysis (PCA) Dan K-Nearest Neighbors (KNN).

I E Hasym, I Susilawati (2019), Dalam penelitian kali ini saya menggunakan 3 jenis ikan cupang untuk melakukan klasifikasi dengan metode PCA(Principal Component Analysis) dan KNN(K-Nearest Neighbors), semoga dengan adanya penelitian ini akan membantu pecinta ikan cupang dalam menentukan jenis ikan cupang. Penelitian ini dimulai dengan pengambilan sampel 3 jenis ikan cupang. Kemudian dilakukan cropping citra untuk menuju tahap proses selanjutnya yakni ekstraksi ciri, training dan testing. Masing-masing ikan di ambil 30 data citra .Total data pelatihan 45 data citra, dan 45 data citra digunakan sebagai data uji, total keseluruhan data 90 data citra. Pada proses ekstraksi ciri menggunakan ekstraksi ciri RGB(Red Green Blue), HSV(Hue Saturation Value) dan area, Proses training dan testing menggunakan algoritma PCA dan klasifikasi menggunakan KNN. Hasil evaluasi pengenalan pola pada citra ikan cupang menggunakan klasifikasi K-NN berdasarkan ekstraksi ciri dengan PCA menghasilkan akurasi sebesar 93,33%.

Dalam jurnal yang berjudul Model Deteksi Kandungan Formalin Pada Ikan Dengan Citra Hue Saturation Values (Hsv) Menggunakan K-Nearest Neighbor. Yufika Agustyan, dkk. (2020), Tujuan dari penelitian ini adalah merancang dan membangun model yang mampu mendeteksi kandungan formalin pada ikan melalui citra ikan tersebut. Metode yang digunakan dalam melaksanakan penelitian ini menggunakan pengolahan citra, dengan metode klasifikasi menggunakan K-nearest neighbor dan ekstraksi ciri menggunakan warna dan tekstur. Berdasarkan penelitian yang telah dilakukan validasi tertinggi didapatkan sebesar 91,67% pada ekstraksi ciri warna dan tekstur, sedangkan validasi terendah sebesar 67,67% pada ekstraksi ciri tekstur. Jika dihitung validasi berdasarkan bagian citra ikan,

didapatkan validasi tertinggi sebesar 86,67% pada citra bagian insang, sedangkan validasi terendah pada citra ikan utuh dengan persentase 81,67%. Berdasarkan uji coba yang telah dilakukan dapat diambil kesimpulan bahwa deteksi dengan menggunakan ekstraksi ciri warna dan tekstur lebih akurat dibandingkan hanya dengan ekstraksi ciri warna ataupun hanya dengan ekstraksi ciri tekstur. Penelitian yang dilakukan juga menunjukkan bahwa semakin besar nilai K (tetangga) akan semakin kecil terjadinya kesalahan.

Dalam jurnal yang berjudul *Case Based Reasoning Menggunakan Algoritma Knearest Neighbors Untuk Penanganan Penyakit Ikan Cupang Hias*. Ariyono Setiawan, dkk. (2020), Tujuan yang hendak dicapai adalah merancang suatu prototipe sistem penalaran berbasis kasus (CBR) yang dapat digunakan sebagai alat bantu untuk melakukan diagnosa penyakit pada ikan hias. Metode yang digunakan dalam penelitian ini menggunakan algoritma K Nearest Neighbors. Kesimpulan yang diperoleh adalah prototipe sistem CBR ini dapat mendiagnosa penyakit ikan hias dengan baik dan hasil analisisnya dapat diterima oleh pakar ikan Cupang hias maupun pengguna non pakar. Namun sistem ini masih memiliki kekurangan yaitu keterbatasan dalam basis pengetahuannya. Penggunaan sistem CBR ini tidak dimaksudkan untuk menggantikan pakar perikanan melainkan sebagai alat bantu dalam mendiagnosa penyakit pada ikan Cupang hias.

Dalam jurnal yang berjudul *Klasifikasi Citra Tekstur Menggunakan K-Nearest Neighbour Berdasarkan Ekstraksi Ciri Metode Matriks Kookurensi*. Asri Junita Arriawati, dkk. (2010), Peneliti ini bertujuan untuk mengetahui bahwa citra-citra tekstur dari basis data Vistex yang diujikan memiliki nilai-nilai koefisien matriks

kookurensi sebagai ciri tekstur yang bervariasi satu dengan yang lain. Ciri-ciri tersebut mempengaruhi hasil klasifikasi sehingga dapat menyebabkan kesalahan klasifikasi. Hasil pengujian terhadap citra uji dari citra belajar menghasilkan tingkat pengenalan sebesar 100% untuk nilai $k = 1$ dan hasil pengujian terhadap citra uji di luar citra belajar menghasilkan tingkat pengenalan sebesar 55,557% untuk nilai $k = 3$.

Dalam jurnal yang berjudul Penerapan Content Based Image Retrieval Untuk Pengenalan Jenis Ikan Koi. Nazil Rizqi Hanggara, dkk (2021), Penelitian ini bertujuan Membangun sistem pengenalan jenis berdasarkan citra bisa dijadikan solusi dari permasalahan. ContentBased Image Retrieval (CBIR) merupakan suatu metode yang dikembangkan dari image retrieval untuk mencari gambar dari suatu database gambar yang besar. Gambar Koi pada CBIR akan dilakukan perhitungan dengan Convolutional Neural Network untuk mendapatkan fitur dari gambar kemudian dilakukan perbandingan kemiripan dengan algoritma Euclidean Distance untuk mendapatkan Hasil dari aplikasi pengenalan jenis ikan Koi ini yaitu urutan citra dari database Koi dimulai dari yang memiliki tingkat kemiripan paling tinggi dengan gambar query. Penelitian ini menggunakan 80 citra dataset dari 4 jenis ikan Koi. Pengujian dari penelitian ini dilihat dari kemampuan membedakan citra koi atau bukan dan dihitung dari persentase data sesuai yang di retrieval dengan hasil yang didapat yaitu mencapai 65%.

Dalam jurnal yang berjudul Ekstraksi Fitur Warna, Tekstur dan Bentuk untuk Clustered-Based Retrieval of Images (CLUE). I Gusti Rai Agung Sugiarta. dkk, (2017), Penelitian ini berfokus pada penyiapan fitur dari sebuah gambar

berdasarkan warna, tekstur, dan bentuk. Fitur warna menggunakan metode Color Histogram, fitur tekstur menggunakan Gray Level Occurance Matrix (GLCM), dan bentuk menggunakan Edge Direction metode Canny Edge Detection. Teknik ekstraksi warna, tekstur, dan bentuk menghasilkan 18 (delapan belas) buah fitur yang mampu digunakan sebagai fitur di proses Clustering gambar. Pengukuran pencarian gambar menunjukkan penggunaan fitur bentuk memiliki nilai pengukuran tertinggi yaitu recall sebesar 1 dan precision sebesar 0,44.

Dalam jurnal yang berjudul Identifikasi Daging Sapi Dan Daging Babi Menggunakan Fitur Ekstraksi Grey Level Co-Occurrence Matrix Dan K-Nearest Neighbor Classifier. Ferry Anggriawan Susanto, Dkk. (2015), Peneliti ini bertujuan mengidentifikasi daging sapi dan daging babi dengan menggunakan fitur ekstraksi Gray level co-occurrence matrix (GLCM) dengan metode klasifikasi K-Nearest Neighbor (K-NN) Classifier. Tahap penelitian dimulai dengan pengumpulan citra daging sapi dan daging babi. Selanjutnya akan diekstraksi fitur dengan menggunakan fitur GLCM. Tahap berikutnya adalah melakukan klasifikasi antara citra latih dan citra uji dengan menggunakan KNearest Neighbor (K-NN) Classifier. Hasil dari klasifikasi tersebut akan dihitung tingkat akurasi. Serta Membandingkan hasil akurasi dengan mengubah arah sudut pada GLCM dan Jumlah k pada K-NN Classifier.

Dalam jurnal yang berjudul Ekstraksi Citra menggunakan Metode GLCM dan KNN untuk Indentifikasi Jenis Anggrek (Orchidaceae). Danar Putra Pamungkas. (2019), Secara umum jenis-jenis Anggrek memiliki kesamaan warna, tekstur dan kelopak bunga, hal ini menyebabkan seseorang mengalami kesulitan dalam

mengidentifikasi jenis anggrek, terutama masyarakat yang belum mengetahuinya ciri-ciri beberapa jenis anggrek. Oleh karena itu proses mengidentifikasi anggrek perlu dilakukan secara otomatis dengan sistem komputer agar diharapkan dapat mempermudah dalam mengidentifikasi jenis-jenis anggrek. Dalam penelitian ini menggunakan metode GLCM untuk ekstraksi ciri dan metode KNN untuk proses mengidentifikasi anggrek atau orchidaceae. Tahapan identifikasi citra anggrek adalah mengubah ukuran awal gambar, konversi ke derajat abu-abu, median filter, ekstraksi fitur metode GLCM dan identifikasi dengan KNN metode. Tingkat keberhasilan identifikasi Orchidaceae atau anggrek mencapai 80% dengan rata-rata 77%. Nilai K mempengaruhi tingkat keberhasilan identifikasi, semakin besar nilai K semakin kecil akurasi.

Dalam jurnal yang berjudul Klasifikasi Ikan Menggunakan Oriented Fast And Rotated Brief (Orb) Dan K-Nearest Neighbor (Knn). Mirza Ramadhani dan Darlis Heru Murti, (2018), Peneliti ini mengusulkan metode untuk mendeteksi dan mengenali jenis objek ikan menggunakan metode ORB dan KNN. Pengaplikasian dari metode ORB diterapkan untuk ekstraksi fitur dari gambar yang diambil. Kemudian hasil tersebut akan diklasifikasi menggunakan KNN untuk menentukan label kelas yang tepat dari input data ikan. ORB memiliki dua bagian utama yaitu FAST keypoint detector dan BRIEF rotation yang disempurnakan sedemikian rupa agar dapat meningkatkan kemampuannya dalam mendeteksi dan mendeskripsi fitur, misalnya pada fitur FAST yang mempunyai komputasi rendah yang kemudian dimodifikasi lagi untuk menambahkan orientasi centroid sehingga dapat mendeteksi objek secara akurat dengan waktu yang lebih singkat, dan fitur BRIEF

yang didesain ulang agar dapat mendeskripsi fitur pada citra sehingga tahan terhadap perubahan transformasi/deformasi objek. Hasil uji coba menunjukkan bahwa metode yang diusulkan pada penelitian ini mencapai akurasi klasifikasi sebesar 97,5%..

Dalam jurnal yang berjudul Analisis Matthew Correlation Coefficient pada K-Nearest Neighbor dalam Klasifikasi Ikan Hias. Novia Hasdyna , Rozzi Kesumadinata. (2020), Peneliti ini bertujuan untuk menunjukkan bahwa nilai matthew correlation coefficient pada k-nn dengan menggunakan Euclidean distance diperoleh nilai MCC tertinggi pada class medium sebesar 0,786542. Nilai mcc tertinggi kedua pada class premium senilai 0,567434. Nilai mcc terendah adalah pada class low sebesar 0,435269. Dengan demikian, dapat dikatakan nilai korelasi matthew.

2.2 Matriks Penelitian

Rangkuman dari penelitian terdahulu yang telah dijelaskan dalam sub sebelumnya diringkas ke dalam tabel 2.1 berikut ini:

Tabel 2.1 Matriks Penelitian

NO.	Penulis	Fitur			Objek										Program			Metode								
		Tekstur	Bentuk	Warna Nilai	Kain	Ikan Koi	coffe	Ikan Cupang	Sapi dan Babi	Nilai UKT	Bunga	etara tekstur	Mamogram	Ikan bias	Leopard Gecko	Matlab	WEB	Android	CLUE	Hybrid Deep Learning	Support Vector Machine	KNN	RGB	Sobel Edge Detection	Naive Bayes	GLCM
1	Tri Adhi Atmaji & Catur Supriyanto (2019)			V	V										V						V			V		V
2	Arif Wicaksono (2017)	V		V	V										V						V					
3	Amadea Sapphira, dkk (2012)		V			V									V					V						
4	Milfa Yetri, dkk (2020)	V	V		V										V								V			
5	Anifuddin Azis (2020)	V				V										V		V								
6	Novia Hasdyn & Rozzi Kesuma dinata (2020)	V	V										V		V						V					
7	IE Hasy & I Susilawati(2019)		V		V			V							V						V	V				
8	Yufika Agustyani, dkk (2020)	V	V					V							V						V					
9	Ariyono Setiawan, dkk (2020)		V					V							V						V					
10	Asri Junita, dkk (2011)	V										V			V						V					
11	Nazil Rizqi Hanggara (2011)		V		V										V											V
12	I Gusti Rai Agung Sugiarta, dkk (2017)	V	V		V										V			V								
13	Ferry Angeriawan Susanto, dkk (2015)	V						V							V						V					
14	Danar Putra Pamungkas(2019)	V	V		V								V		V						V					V
15	Mirza Ramadhani & Darlis Heru Murti (2018)	V	V		V										V	V					V					
16	Novia Hasdyna (2020)	V													V	V					V					
Penelitian yang diusulkan					V	V									V						V	V				

1.3 Penelitian Terdekat

Penelitian terdekat ini menjadi salah satu acuan dalam melakukan penelitian sehingga dapat membandingkan metode yang digunakan dalam penelitian yang dilakukan. Penelitian terdekat diambil dari penelitian terdahulu yang paling mendekati dengan penelitian yang akan dilakukan. Penelitian terdekat ini dijadikan perbandingan dengan penelitian yang akan dilakukan sehingga dapat diketahui perbedaan apa saja yang ada pada penelitian ini.

Dari penelitian (Tri Adhi Atmaji & Catur Supriyanto, 2019), Ada kemiripan dengan tujuan penelitian yang akan dilakukan dimana teknologi yang akan digunakannya juga sama. Yang berbeda adalah dari segi metode penelitiannya. (Arif Wicaksono 2017), memiliki kesamaan dari segi metode yang akan digunakan yaitu metode GLCM. Akan tetapi ada algoritma tambahan yaitu NAIVE BAYES. (Amadea Sapphira, dkk 2011), memiliki tingkat kesamaan, karena penelitian ini akan mengklasifikasikan jenis ikan koi berdasarkan gambar Menggunakan Zero Parameter Simple Linear Iterative Clustering dan Support Vector Machine

2.4 Citra Digital

Citra Digital Citra digital dapat diwakili oleh sebuah matriks dua dimensi $f(x,y)$ yang terdiri dari kolom (M) dan baris (N), dimana perpotongan antara kolom dan baris disebut piksel (pixel) disebut juga picture element atau elemen terkecil dari suatu citra. Citra merupakan fungsi kontinu (continue) dengan intensitas cahaya pada bidang dua dimensi. Agar suatu citra dapat diolah dengan komputer digital, maka citra tersebut harus dipresentasikan secara numerik dengan nilai-nilai diskrit. Representasi dari fungsi kontinyu menjadi nilai-nilai diskrit disebut digitalisasi citra. Pengolahan citra digital adalah suatu pemrosesan gambar 2 dimensi dengan menggunakan komputer sebagai alat bantu untuk mengolah sebuah citra 2 dimensi. Secara garis besar pengolahan citra sendiri adalah mengacu pada pemrosesan setiap data 2 dimensi. Citra digital sendiri adalah

merupakan sebuah larik (array) yang berisikan sebuah nilai-nilai real maupun kompleks yang direpresentasikan dengan sebuah deretan bit tertentu. Sebuah citra dapat diartikan sebagai fungsi $f(x,y)$ yang berukuran M baris dan N kolom, dengan x dan y adalah koordinat spasial, dan amplitude f di titik koordinat (x,y) yang dinamakan intensitas atau tingkat keabuan dari citra pada sebuah titik yang ditunjukkan.

2.5 Citra RGB

Dalam sebuah citra digital terdiri dari tiga bidang citra yang saling lepas, dengan masing-masing bidang citra yang terdiri dari tiga buah warna utama, yaitu warna merah, warna hijau, dan warna biru. Setiap titik pada layar menunjuk pada piksel yang memiliki tiga buah kombinasi angka, yaitu R (red), G (green), dan B (blue) yang menentukan posisi dari warna merah, warna hijau, dan warna biru. Maka dari itu dalam suatu piksel akan mewakili tiga byte memori yang masing-masing terdiri dari 1 byte untuk warna merah, 1 byte untuk warna hijau, dan 1 byte untuk warna biru. Sehingga dalam sebuah piksel yang menunjuk pada sebuah nilai intensitas kecerahan suatu citra dapat dituliskan sebagai contoh misalkan dengan warna RGB(25,70,255), sedangkan warna putih bisa dituliskan dengan RGB(255,255,255), dan hitam dengan RGB(0,0,0).

2.6 Citra Grayscale

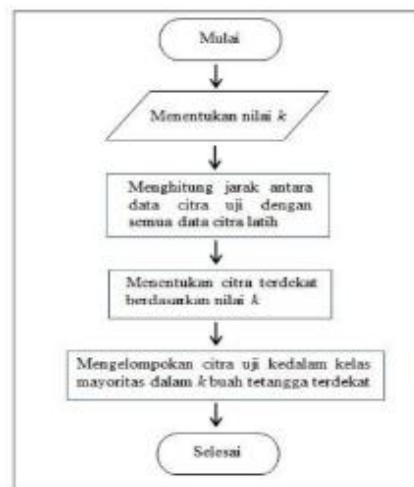
Citra grayscale adalah merupakan citra digital yang hanya memiliki satu nilai kanal atau hanya memiliki nilai tunggal pada setiap pikselnya, dengan kata lain satu nilai kanal pada piksel tersebut merupakan nilai bagian dari RGB. Nilai tersebut menunjukkan besarnya tingkat intensitas pada suatu citra grayscale. Warna yang dimiliki pada citra grayscale adalah terdiri dari warna hitam, keabuan, dan putih.

2.7 K-Nearest Neighbor (K-NN)

Klasifikasi merupakan proses yang menggambarkan dan membedakan kelas data atau konsep dengan tujuan agar data baru dapat diprediksi atau agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui. Algoritma klasifikasi saat ini sangat sering digunakan secara luas, salah satu algoritma klasifikasi adalah KNN (K-Nearest Neighbor). Dalam algoritma klasifikasi KNN termasuk dalam kelompok instance-based learning. Cara kerja algoritma KNN dalam mengklasifikasi data baru adalah dengan mencari kelompok k objek dalam data training yang paling dekat atau mirip dengan data baru atau data testing. Algoritma KNN adalah termasuk dalam algoritma supervised learning, dimana hasil dari query instance akan diklasifikasikan berdasarkan label terbanyak dari KNN atau berdasarkan data training . Algoritma KNN bekerja hanya berdasarkan jarak terpendek dari query instance ke data sample, setelah itu mengurutkan data yang mempunyai jarak terpendek dan kemudian diambil k buah data terdekat. Kelas yang paling banyak ditemui pada k buah tetangga terdekat, maka data uji termasuk dalam kategori tersebut. Ada banyak cara untuk mengukur jarak antara data baru dan data lama (data training). Perhitungan jarak antara data query instance dan data uji dapat dihitung menggunakan Euclidean distance. Langkah dalam melakukan klasifikasi dengan algoritma KNN adalah sebagai berikut:

1. Langkah pertama adalah dengan menentukan k, yaitu jumlah tetangga terdekat. Parameter k bernilai ganjil, misal 1,3,5,7,9.
2. Kemudian menghitung jarak antara citra latih dengan citra uji. Ada banyak cara untuk mengukur jarak antara citra latih dengan citra uji. Salah satunya menggunakan perhitungan jarak Euclidean distance.

3. Kemudian mengurutkan hasil pencarian jarak terdekat, setelah itu tetapkan tetangga terdekat berdasarkan k yang ditentukan.
4. Periksa label pada k buah data terdekat.
5. Kemudian kelompokkan citra uji kedalam kelas mayoritas dari k buah tetangga terdekat berdasarkan data citra latih. Berikut adalah alur diagram dari proses klasifikasi dengan menggunakan algoritma klasifikasi KNN :



Gambar 2.1 Diagram alir klasifikasi