

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Landasan Teori**

Diperlukan kajian teoritis dan studi pendahuluan terlebih dahulu pada objek untuk dapat memilih variabel apa yang akan diteliti. (Prof. Dr. Sugiyono 2020). Dalam penelitian ini akan dilakukan kajian pustaka yang menunjang penelitian diantaranya terkait ekstraksi citra, GLCM, dan algoritma *Weight-KNN*. Literatur bersumber dari jurnal–jurnal dan buku yang sumbernya terpercaya.

##### **2.1.1 Ikan Gabus Hias**

Ikan gabus memiliki tampilan yang eksotis dan menarik sehingga banyak dimanfaatkan sebagai ikan hias. Selain itu, ikan gabus merupakan salah satu ikan air tawar yang hidup di kawasan Asia Tenggara yang tergolong ke dalam ikan karnivora. Dikenal juga sebagai ikan yang dapat konsumsi dan diperjualbelikan di pasaran dalam ukuran kecil. (Listyanto dan Andriyanto 2009).

##### **2.1.2 *Machine Learning***

*Machine Learning* (ML) atau pembelajaran mesin adalah bidang penelitian penting dari kecerdasan buatan yang dapat membantu komputer dalam melakukan pemodelan berdasarkan pengalaman yang secara akurat memprediksi peristiwa masa depan (Dogan dan Birant 2021). *Machine learning* dapat menangani data dalam jumlah besar dengan cara yang cerdas sehingga dapat memberikan hasil

yang akurat dan tepat (Abijono, Santoso, dan Anggreini 2021). *Supervised learning* teknik dalam ML yang menggunakan data yang telah diberikan label (data latih) untuk melakukan proses pembelajaran dalam mesin, sehingga mesin mampu melakukan klasifikasi ataupun prediksi. Sedangkan *unsupervised learning* tidak mendapatkan data latih karena algoritma *unsupervised learning* tidak bersifat prediktif, sehingga prosesnya membutuhkan pembelajaran dari data yang telah ada sebelumnya (Retnoningsih and Pramudita 2020). Data yang baik didapatkan dari proses ekstraksi fitur, validasi, integrasi dan transformasi serta reduksi ukuran data dan diskretisasi data agar sesuai dengan data latih yang telah disiapkan (Rolansa, Yunita, dan Suheri 2020).

### 2.1.3 Klasifikasi

Klasifikasi adalah sebuah proses untuk menganalisis banyak data yang dikumpulkan menjadi beberapa kelompok sesuai aturan pengelompokan yang dapat digunakan untuk mengklasifikasikan data yang baru (Sanjaya, Setyati, dan Budianto 2020). Berikut adalah proses klasifikasi didasarkan pada empat komponen (Rohman dan Rochcham 2019):

1. *Kelas*, Komponen kelas merupakan variabel dependen yang merepresentasikan '*label*' pada objek dan memiliki karakteristik kategorikal.
2. *Predictor*. Komponen *predictor* merupakan variabel independen yang direpresentasikan oleh karakteristik (atribut) data.
3. *Training dataset*. Komponen ini merupakan satu set data yang berisi nilai dari komponen kelas yang akan diklasifikasikan sesuai komponen *predictor*.

4. *Testing dataset*. Komponen yang berisi data baru untuk diklasifikasikan oleh model yang telah dibuat dan akurasi klasifikasinya akan dievaluasi.

#### 2.1.4 Algoritma K-Nearest Neighbour

*K-Nearest Neighbour* (KNN) merupakan teknik klasifikasi yang simpel, tetapi memiliki hasil klasifikasi yang cukup bagus (Sibarani, 2015). Salah satu algoritma klasifikasi *supervised learning* yang mengelompokkan suatu objek ke dalam data kelas baru berdasarkan data latih dan atribut kelas (Pratama, Mustajib, dan Nugroho 2020). Bekerja dengan cara memilih tetangga terdekat dari data latih (atau di sebut  $k$ ), kemudian hasil klasifikasi ditentukan berdasarkan nilai jarak terdekat atau nilai jarak terkecil yang dihitung menggunakan rumus *euclidean* (Nafiah 2019). Oleh karena itu nilai  $K$  dalam KNN perlu diketahui masing-masing akurasi untuk mendapatkan hasil akurasi yang optimal (Banjarsari, Budiman, dan Farmadi 2015).

Rumus jarak *euclidean* akan digunakan untuk mengetahui jarak antara dua objek  $x$  dan  $y$ , dimana jarak skalar dari kedua data  $x$  dan  $y$  akan dinyatakan oleh matriks  $D(x, y)$ , matriks dengan ukuran  $d$  dimensi (Yofianto, 2010). Algoritma ini hanya melakukan penyimpanan data-data fitur dan klasifikasi data training sampel pada fase *training*. Fitur-fitur yang sama dihitung untuk *testing* data pada fase klasifikasi. Jarak dari data baru terhadap seluruh data *training* sampel dihitung secara berulang namun hanya data sejumlah  $k$  yang paling dekat yang diambil untuk proses penentuan kelas. Data terbanyak dari hasil klasifikasi akan menjadi kelas prediksi dari data yang dihitung.

### 2.1.5 Algoritma Weight-KNN

KNN konvensional menggunakan teknik *vote majority* atau data terdekat terbanyak untuk penentuan kelas bagi data uji dengan ketentuan setiap nilai tetangga terdekat mempunyai nilai *weight* atau bobot yang sama dalam penentuan kelas (Lidya, Sitompul, and Efendi 2015). Metode ini dinamakan metode *Weight-KNN* yang memilih kelas data uji berdasarkan jumlah tertinggi dari rata-rata bobot pada setiap kelas. Maka akan menyebabkan tidak rasionalnya jarak setiap tetangga terdekat terhadap jarak data uji yang dapat mengakibatkan kesalahan dalam proses klasifikasi (Syaliman 2018).

### 2.1.6 Algoritma Gray Level Co-occurrence Matrix

Algoritma *Gray Level Co-occurrence Matrix* (GLCM) merupakan salah satu metode ekstraksi ciri dalam pendekatan citra untuk memperoleh nilai fitur dengan cara menghitung nilai peluang dari hasil perhitungan hubungan terdekat antara 2 piksel dalam jarak dan ukuran sudut tertentu (Musrini, Andriana, dan Hidayat 2017). Selain itu, GLCM juga merupakan salah satu ekstraksi ciri untuk mendapatkan nilai dari masing-masing fitur dengan proses perhitungan pada matriks yang sama yang muncul dalam setiap piksel gambar. Fitur-fitur yang ada dalam GLCM diantaranya adalah energi, entropi, homogenitas, kontras, korelasi, mean dan dissimilariti. GLCM dapat digunakan untuk ekstraksi sebuah gambar dikarenakan setiap gambar memiliki tekstur dan pola yang berbeda-beda.

Cara menghitung nilai peluang dari hasil perhitungan hubungan terdekat antara 2 piksel dalam jarak dan ukuran sudut tertentu adalah ketika GLCM

berukuran  $L \times L$  dimana  $L$  menyatakan tingkat keabuan, dengan elemen  $P(x_1, x_2)$  yang merupakan distribusi probabilitas dari pasangan piksel yang memiliki tingkat keabuan  $x_1$  pada koordinat  $(j, k)$  dan  $x_2$  pada koordinat  $(m, n)$ .  $P(x_1, x_2)$ . Koordinat pasangan piksel tersebut memiliki jarak  $d$  dan orientasi sudut  $\theta$ . Jarak direpresentasikan dalam piksel dan sudut direpresentasikan dalam derajat. Orientasi sudut terbentuk berdasarkan empat arah sudut yaitu,  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$  dan  $135^\circ$ , jarak antar piksel sebesar 1-piksel (Widodo, Widodo, dan Supriyanto 2018).

### 2.1.7 Confusion Matrix

*Confusion matrix* merupakan suatu metode yang biasa digunakan untuk mengukur akurasi pada data mining. *Confusion matrix* meringkas kinerja klasifikasi *clasifier* sehubungan dengan beberapa data pengujian (Ting, 2010). Matriks ini merupakan matriks dua dimensi berupa tabel matriks, jika data set hanya terdiri dari dua kelas, kelas yang satu dianggap sebagai positif dan yang lainnya negatif.

Tabel 2. 1 Tabel *Confusion Matrix 2x2*

Klasifikasi yang benar	Diklasifikasikan sebagai	
	+	-
+	<i>True Positive</i>	<i>False Negative</i>
-	<i>False Positive</i>	<i>True Negative</i>

Keterangan:

1. *True positive* : jumlah record positif yang diklasifikasikan sebagai positif,
2. *False positive* : jumlah record negatif yang diklasifikasikan sebagai positif,
3. *False negative* : jumlah record positif yang diklasifikasikan sebagai negatif,
4. *True negative* : jumlah record negatif yang diklasifikasikan sebagai negatif.

## 2.2 Penelitian Terkait dan Kebaruan Penelitian

Tantangan penelitian dijawab dengan memperluas *state-of-the-art* bidang penelitian dengan memasukkan unsur-unsur tambahan yang dapat memenuhi kebutuhan klasifikasi ikan gabus hias dengan metode pendekatan citra menggunakan GLCM dan *Weight-KNN*. Berikut ini adalah perbandingan hasil penelitian yang terkait dengan kontribusi terhadap proses klasifikasi citra (*state-of-the-art* penelitian terkait), identifikasi kebaruan penelitian, dan relevansi penelitian yang dilakukan.

Tabel 2. 2 *State of The Art* Penelitian Terkait

No	Nama Pengarang	Tahun	Judul	Masalah	Solusi
1.	M. Ramadhani, D. Murti	2018	Klasifikasi Ikan Menggunakan <i>Oriented Fast and Rotated Brief</i> (Orb) Dan <i>K-Nearest Neighbor</i> (KNN)	Diperlukan adanya metode yang tepat untuk dapat mengklasifikasi ikan yang invariant terhadap berbagai macam perubahan	Menggunakan metode ORB dan KNN untuk proses klasifikasi karena tahan terhadap perubahan seperti ukuran, pencahayaan dengan akurasi 97,50%.
2.	A.Pariyandani dkk	2019	Klasifikasi Citra Ikan Berformalin Menggunakan Metode KNN dan GLCM	Diperlukan sistem yang mampu mendeteksi ikan yang berformalin secara otomatis.	Deteksi ikan berformalin dengan menggunakan algoritma GLCM dan KNN dengan tingkat akurasi 92%.
4.	D. Pamungkas	2019	Ekstraksi Citra menggunakan Metode GLCM dan KNN untuk Identifikasi Jenis Anggrek ( <i>Orchidaceae</i> )	Tidak adanya sistem atau aplikasi yang dapat mengidentifikasi jenis bunga anggrek otomatis.	Melakukan deteksi Bunga Anggrek melalui pendekatan Ekstraksi Citra menggunakan Metode GLCM dan KNN yang mampu mencapai akurasi 77%.
4.	M. Akbar dkk	2021	Klasifikasi Jenis Ikan Cupang Menggunakan Metode GLCM Dan KNN.	Belum adanya sistem untuk menentukan jenis ikan cupang,	Sistem klasifikasi jenis ikan cupang menggunakan GLCM dan KNN dengan akurasi 80%.
5.	R. Paweing dkk	2016	Ekstraksi Fitur Berdasarkan Deskriptor Bentuk dan Titik Salien Untuk Klasifikasi Citra Ikan Tuna	Sistem pemisahan manual yang sering terjadi adalah kesalahan pemisahan dan waktu pemisahan yang tidak stabil.	Klasifikasi ikan Tuna dengan pendekatan Ekstraksi Fitur Berdasarkan Deskriptor Bentuk dan Titik Salien yang memperoleh akurasi sebesar 73,33%.

Tabel 2. 3 Tabel Kebaruan Penelitian

No	Nama Peneliti	Tahun	Judul	Algoritma		Objek
				Ekstraksi	Klasifikasi	
1	M. Ramadhani, D. Murti	2018	Klasifikasi Ikan Menggunakan <i>Oriented Fast and Rotated Brief (Orb)</i> Dan <i>K-Nearest Neighbor (Knn)</i>	ORB	KNN	Ikan
2	A.Pariyandani dkk	2019	Klasifikasi Citra Ikan Berformalin Menggunakan Metode KNN dan GLCM	GLCM	KNN	Ikan Berformalin
3	D. Pamungkas	2019	Ekstraksi Citra menggunakan Metode GLCM dan KNN untuk Identifikasi Jenis Anggrek ( <i>Orchidaceae</i> )	GLCM	KNN	Anggrek
4	M. Akbar dkk	2021	Klasifikasi Jenis Ikan Cupang Menggunakan Metode GLCM Dan KNN.	GLCM	KNN	Ikan Cupang
5	R. Paweing dkk	2016	Ekstraksi Fitur Berdasarkan Deskriptor Bentuk dan Titik Salien Untuk Klasifikasi Citra Ikan Tuna	ROBUST	KNN	Ikan Tuna
6	Abdi Maliki	2021	Klasifikasi Ikan Channa (Gabus Hias) Dengan Pendekatan Citra Menggunakan Algoritma <i>Gray-Level-Co-Occurance-Matrix</i> Dan <i>Weight K-Nearest Neighbour</i>	GLCM	<i>Weight-KNN</i>	Ikan Gabus Hias

Keterangan:

GLCM = *Gray Level Co-Occurance Matrix*

KNN = *K-Nearest Neighbour*

Tabel 2. 1 Tabel Relevansi Penelitian

<b>Peneliti</b>	<b>M Akbar (2021)</b>	<b>Abdi Maliki (2022)</b>
Judul	Klasifikasi Jenis Ikan Cupang Menggunakan Metode GLCM Dan KNN	Klasifikasi Ikan Channa (Gabus Hias) Dengan Pendekatan Citra Menggunakan Algoritma <i>Gray-Level-Co-Occurance-Matrix</i> Dan <i>Weight K-Nearest Neighbour</i>
Masalah Penelitian	Belum adanya sistem untuk menentukan jenis ikan cupang,	Tidak adanya sistem klasifikasi ikan gabus hias secara otomatis. KNN konvensional tidak melihat bobot dari setiap tetangga terdekat untuk menentukan hasil klasifikasi.
Objek Penelitian	Ikan Cupang	Ikan Gabus Hias
Algoritma	GLCM-KNN	GLCM-Weight KNN