

BAB II

LANDASAN TEORI

2.1 KAJIAN PUSTAKA

Landasan teori berisikan beberapa kajian pustaka dari beberapa penelitian yang dilakukan sebelumnya dengan topik yang sama seperti penelitian yang akan dilakukan oleh peneliti, penelitian – penelitian itu diantaranya adalah penelitian yang dilakukan oleh Sari, Widodo, dan Priambodo yang dipublikasikan tahun 2019 dengan judul “Klasifikasi Jenis Citra Makanan Menggunakan *Color Histogram* dan *Level Occurrence Matrix* dengan *K-Nearest Neighbour*” yang membahas tentang cara klasifikasi citra makanan yang dilakukan dengan cara ekstraksi fitur histogram yang ada dalam citra lalu memanfaatkan algoritma *K-Nearest Neighbour* untuk melakukan klasifikasi, dari penelitian tersebut didapatkan bahwa dalam melakukan pengklasifikasian makanan menggunakan ekstraksi fitur *color histogram* memiliki perbedaan hasil dengan menggunakan *Gray Level Co-occurrence Matrix* dengan nilai perbedaan mencapai 20% dimana hasil juga dipengaruhi oleh metode penghitungan jarak yang digunakan [4].

Penelitian selanjutnya tentang pengolahan citra adalah penelitian yang dilakukan oleh Pawening, Kusuma, dan Dijaya yang dipublikasikan tahun 2017 dengan judul “Otomatisasi Klasifikasi Kematangan Buah Mengkudu Berdasarkan Warna dan Tekstur” penelitian ini mempelajari tentang cara pengklasifikasian buah mengkudu berdasarkan tingkat kematangan dengan melihat warna buah mengkudu sebagai salah satu faktor pembeda yang digunakan untuk melakukan klasifikasi mengkudu matang dan mentah , penelitian ini menggunakan klasifikasi fitur warna memiliki keakuratan sebesar 87.00% [5].

Penelitian yang dilakukan oleh Sanjaya, Gusti, dan Pura yang dipublikasikan tahun 2019 dengan judul “*K-Nearest Neighbor for Classification of Tomato Maturity Level Based on Hue, Saturation, and Value Colors*” penelitian ini membahas tentang pengklasifikasian tingkat kematangan buah tomat menggunakan citra buah tomat yang didasarkan nilai warna ,*hue* dan saturasi dengan hasil penelitian bahwa nilai ketetangaan *K-Nearest Neighbor* untuk klasifikasi buah tomat memiliki pengaruh yang tidak terlalu signifikan terhadap hasil klasifikasi, sedangkan nilai ukuran citra memiliki pengaruh yang signifikan terhadap hasil klasifikasi dengan akurasi tertinggi mencapai 92,5% pada citra buah tomat berukuran 1000x1000 pixel dengan nilai ketetangaan 3 [6].

Penelitian oleh Liantoni dan Annisa pada tahun 2018 dengan judul “*Fuzzy K-Nearest Neighbor* Pada Klasifikasi Kematangan Cabai Berdasarkan Fitur *HSV* Citra” meneliti cara klasifikasi cabai dengan cara melakukan ekstraksi fitur warna *RGB* dan *HSV* yang digabungkan dengan algoritma *Fuzzy K-Nearest Neighbor* mendapatkan hasil kematangan cabai rawit yang dilakukan pada 4 kelas kematangan yaitu kelas cabai mentah, cabai setengah matang, cabai matang, dan

cabai busuk menggunakan *K-NN* dengan fitur *HSV* memiliki akurasi sistem sebesar 86,66% [7].

Penelitian yang dilakukan oleh Sari, Febriani, dan Wihandika yang dipublikasikan tahun 2019 dengan judul “Klasifikasi Citra Kue Tradisional Indonesia Berdasarkan Ekstraksi Fitur Warna *RGB Color Moment* Menggunakan *K-Nearest Neighbor*” membahas tentang klasifikasi citra kue tradisional yang dilakukan dengan melakukan ekstraksi fitur warna dalam citra dengan hasil pengujian yang dilakukan terhadap 58 data uji mendapatkan hasil bahwa dengan melakukan perubahan pada nilai K atau nilai total tetangga yang digunakan akan mempengaruhi hasil evaluasi secara signifikan [8].

2.2 CABAI

Cabai (*Capsicum sp*) adalah jenis tanaman perdu yang termasuk dalam family terong – terongan (*Solanaceae*) yang dikenal sebagai salah satu bumbu masakan yang dimanfaatkan untuk memberikan sensasi terbakar di mulut, tanaman cabai berasal dari benua amerika yang digunakan oleh suku asli benua amerika seperti suku *Maya* (Amerika Tengah), suku *Inca* (Amerika Selatan), dan suku *Aztek* (Meksiko), cabai dibawa oleh pelaut dan pedagang legendaris *Cristoper Columbus* yang membawa cabai ke dunia baru (Eropa) melalui perdagangan ke Italia, Spanyol, hingga kemudian ke seluruh belahan dunia pada tahun 1492 [9]. Contoh cabai dari amerika latin yang dibawa ke seluruh dunia adalah *capcicum annum* yang dapat dilihat di gambar 2.1.



Gambar 2. 1 *Capcicum Annum* Salah Satu Varietas Cabai [9]

Tanaman cabai dibudidayakan karena cabai merupakan komoditas dengan nilai pasar yang tinggi, cabai adalah salah satu komoditas sayuran populer karena fungsi utamanya sebagai pemberi rasa pedas, penguat rasa dan bumbu masakan. Selain karena dijual sebagai bumbu masakan cabai juga bermanfaat untuk manusia karena pada cabai terdapat kandungan antioksidan yang sangat baik untuk melindungi tubuh dari radikal bebas, cabai juga memiliki kandungan

vitamin c yang cukup tinggi sehingga baik dikonsumsi oleh tubuh dengan kadar yang wajar supaya tidak menimbulkan gangguan pada sistem pencernaan.

Cabai memiliki sebutan yang berbeda-beda di Indonesia karena cabai dibudidayakan secara luas di Indonesia, contohnya seperti Lombok (Jawa), cabe dan cengek (Sunda), lasina (Batak Karo), campli (Aceh), lado (Minangkabau), dan risak (Papua). Karena cabai memiliki daerah konsumsi yang luas petani cabai akan melakukan panen cabai dalam berbagai tingkat kematangan agar dapat didistribusikan secara baik dan tiba ke daerah tujuan dalam keadaan segar dan tidak busuk, Panen tanaman di dataran rendah dilakukan ketika umur cabai mencapai 60 hingga 75 hari pasca tanam dimana cabai yang dijual ke industri pengolahan cabai berupa cabai matang berwarna merah 100%, sementara cabai yang digunakan untuk konsumsi pasar dipanen ketika masih dalam kondisi matang 80% untuk menghindari kebusukan di dalam pengiriman dan dalam kasus cabai yang akan dikirim keluar pulau menggunakan kapal, cabai akan dipanen dalam kondisi masih hijau (belum matang) untuk menghindari kebusukan karena jarak yang jauh.

2.3 SMART FARMING

Smart Farming dapat didefinisikan sebagai tata cara penggunaan dan pemanfaatan variabilitas spasial dan terporal memanfaatkan teknologi informasi dengan tujuan untuk melakukan peningkatan efektifitas produksi pertanian. Pertanian didasarkan pada keakuratan penggunaan *input* produksi, sehingga dapat diperoleh keuntungan penghematan biaya tenaga kerja dan mendapat hasil panen yang baik [10].

Jenis-jenis teknologi yang diaplikasikan dalam *Smart Farming* dapat dipisahkan menjadi tiga kategori utama berdasarkan fungsi utama teknologi tersebut seperti:

1. Teknologi yang berfokus pada akuisisi data seperti teknologi survei, teknologi navigasi, teknologi pemetaan, dan pengukuran serta penginderaan yang digunakan di lahan pertanian.
2. Teknologi analisis dan evaluasi data berisi teknologi pemilihan keputusan sederhana berdasarkan hal di lapangan seperti cuaca dan curah hujan hingga manajemen pertanian kompleks yang memanfaatkan sistem informasi dengan banyak variabel yang berbeda-beda.
3. Teknologi aplikasi yang berisi aplikasi pengolahan maupun inventori dengan fokus pada pengaplikasian tingkat variabel serta teknologi panduan pertanian.

Smart Farming memanfaatkan teknologi-teknologi dari berbagai macam disiplin ilmu dengan harapan untuk mendapatkan hasil pertanian yang maksimal dan efisien seperti:

1. Pemanfaatan lahan yang efektif dan efisien menggunakan teknologi pemetaan.
2. Penggunaan mesin otomatis seperti *sprinkler* dan *conveyor belt*.

3. Akuisisi data lingkungan berbasis citra menggunakan berbagai jenis kamera.
4. Akuisisi data berbasis sensor menggunakan sensor – sensor yang dipasang di sekitar lahan pertanian.
5. Penggunaan komputer untuk melakukan pengolahan data yang didapatkan dan melakukan perencanaan pengelolaan pertanian.
6. Penggunaan internet untuk mengetahui informasi iklim dan cuaca daerah sekitar pertanian.
7. Penggunaan media sosial untuk mengetahui minat pasar dan proyeksi kebutuhan produk pertanian.

Kategori teknologi pertanian cerdas yang dibagi berdasarkan kategorinya dan contoh – contohnya ditunjukkan dalam tabel 2.1.

Tabel 2. 1 Daftar Kategori Teknologi Pertanian Cerdas dan Contohnya [11]

Teknologi Cerdas Pertanian	Kategori Teknologi	Contoh Sistem
Teknologi akuisisi data	GNSS	Sistem navigasi global satelit (GNSS)
		Real Time Kinematic
		<i>GNSS Difrential</i>
		<i>Pinpoint GNSS</i>
	Pemetaan	Pemetaan tanah
		Peta Ketinggian
		Pemetaan hasil pertanian
	Akuisisi data lingkungan berbasis citra	Kamera <i>LIDAR</i>
		Kamera <i>RGB</i>
		Kamera <i>Thermal</i>
	Akuisisi data lingkungan berbasis sensor	Sensor kelembapan tanah
		<i>Tensiometer</i>
Mesin dan Properti	Sensor pembuangan panas	
	<i>Drone</i>	
	Komputer	
Analisis Data & Evaluasi Teknologi	Analisis berdasar data lapangan	Sprinkler otomatis
		Sistem pendukung pengambilan keputusan
Teknologi Pengaplikasian Sistem Pertanian Presisi	<i>Guidance Technoloy</i>	Sistem pendeteksi perubahan cuaca
		<i>Automatic guidance</i>
	<i>Variable Aplicator</i>	Lalu – lintas Pertanian
		Pengaplikasian pestisida otomatis berdasar variabel tanah
		Pengaplikasian pupuk cair otomatis berdasarkan kelembapan tanah

2.4 PENGOLAHAN CITRA

Citra adalah fungsi intensitas yang berada dalam bidang dua dimensi. Intensitas yang dihasilkan oleh cahaya, dan cahaya adalah salah satu bentuk energi maka fungsi intensitas citra cahaya terletak di antara: $0 < f(x, y) < \infty$. Citra yang dilihat terdiri atas berkas – berkas cahaya yang dipantulkan oleh benda – benda di sekitarnya dan ditangkap oleh reseptor cahaya yang dapat berupa organisme menggunakan mata mereka maupun perangkat keras yang mampu membaca dan mengolah citra seperti kamera [12].

Berdasarkan jenisnya citra dapat dibedakan menjadi dua jenis yaitu citra statis (*still images*) dan citra bergerak (*moving images*). Citra statis adalah sebuah citra tunggal yang tidak berganti – ganti maupun bergerak, citra bergerak adalah rangkaian citra statis yang ditampilkan secara bergantian dan memiliki urutan tertentu dalam rentang waktu singkat hingga dapat memberi kesan seolah-olah citra tersebut mengalami pergerakan, misalnya animasi, video pendek dan gif. Dalam penelitian ini karena penelitian memiliki batasan pada citra statis, selanjutnya citra statis akan disebut sebagai citra. Berdasarkan cara pembentukannya citra dapat dibedakan menjadi 2 yaitu citra diskrit dan citra kontinyu. Citra diskrit adalah citra yang dihasilkan dengan melakukan proses digitalisasi terhadap citra kontinu. Beberapa cara melakukan digitalisasi adalah dengan menggunakan alat seperti scanner dan kamera digital, sedangkan citra kontinu adalah citra yang ditangkap oleh sistem optik dengan masukan sinyal analog, dengan contoh kamera analog dan mata manusia. Citra diskrit sering juga disebut sebagai citra digital karena pembuatannya [13], seperti gambar Lena Sjooblom pada gambar 2.2 yang sering digunakan sebagai *icon* citra statis dalam *image processing*.



Gambar 2. 2 Lena Sjooblom dalam citra statis [14]

Pengolahan citra (*image processing*) adalah proses mengolah citra dengan menggunakan komputer menggunakan beberapa perangkat lunak tertentu agar menjadi citra yang memiliki kualitas yang lebih baik jika dibandingkan dengan citra awal. Secara umum pengolahan citra dilakukan kepada sebuah citra dengan kondisi – kondisi tertentu untuk mendapatkan tujuan seperti:

1. Melakukan perbaikan dan atau memodifikasi citra guna memperbaiki kualitas penampakan atau untuk menonjolkan beberapa aspek – aspek tertentu dalam citra yang memiliki informasi yang ingin diambil.
2. Elemen – elemen yang ada di dalam sebuah citra ingin dilakukan perlakuan tertentu seperti pengelompokan, penyocokkan dan atau pengukuran untuk mendapatkan variabel tertentu dalam citra.
3. Citra tersebut perlu dilakukan penggabungan dengan bagian citra lain atau sebagian citra perlu dilakukan pemotongan untuk mendapat bagian tertentu dari citra awal.

Pada prinsipnya pengolahan citra dilakukan untuk melakukan perbaikan kualitas citra agar citra tersebut dapat mudah diinterpretasikan dan diproses oleh pengguna citra seperti mata manusia maupun mesin seperti komputer, teknik pengolahan citra melakukan transformasi citra awal menjadi citra yang lebih mudah dicerna dari citra awal. Jadi pengolahan citra memiliki masukan berupa citra dan *output* berupa citra yang citra akhir (*output*) memiliki kualitas yang lebih baik daripada citra masukan, contohnya seperti apabila citra awal memiliki beberapa aspek yang kurang baik, seperti citra yang terlalu gelap atau terlalu terang sehingga sulit untuk dicerna dan diproses, citra yang kabur sehingga sulit menentukan variabel dalam, citra dengan warna yang kurang tajam dan kondisi lain yang dapat ditingkatkan kualitas citranya menggunakan pengolahan citra guna mendapatkan hasil citra yang lebih baik, untuk melakukan manipulasi citra diperlukan teknik mengolah citra dengan menggunakan perangkat tertentu. Teknik mengolah citra dinamakan pengolahan citra [13]. Pengolahan citra dilakukan untuk mengubah properti dari sebuah citra statis seperti pada gambar 2.3 dibawah.



Gambar 2. 3 Contoh pengaplikasian pengolahan citra [14]

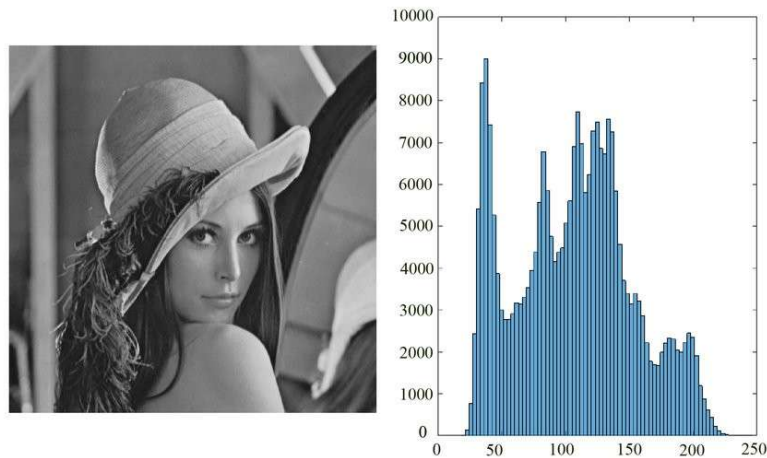
2.5 HISTOGRAM

Histogram adalah sebuah grafik yang mengindikasikan jumlah kemunculan setiap level keabuan pada suatu citra [15]. Dalam sebuah citra variabel histogram bisa dipergunakan guna menyelesaikan suatu masalah yang ada dalam melakukan penelitian pada sebuah citra, pada penelitian ini fitur histogram dipergunakan sebagai pembatas dalam melakukan pengambilan gambar agar variabel yang didapatkan sesuai dengan batasan yang ditentukan, membantu meningkatkan visual penampilan gambar serta dapat digunakan sebagai alat forensik dalam menentukan pemrosesan citra apa yang telah dilakukan pada citra masukan [16].

Histogram dapat diartikan menjadi kerapatan distribusi pixel – pixel yang menghasilkan intensitas yang spesifik dalam suatu citra dalam masing – masing pixel. Histogram adalah frekuensi dari sebuah intensitas, artinya citra yang memiliki kanal lebih dari 1 memiliki histogram representasi untuk masing – masing kanal yang dimiliki, sedangkan apabila citra adalah citra hitam putih (*grayscale*) citra tersebut hanya memiliki satu histogram dengan intensitas dengan rentang $I_{(u,v)} \in [0, K - 1]$ dimana nilai dari K adalah intensitas dari derajat keabuan citra yaitu $K = 2^8 = 256$. Histogram dapat didefinisikan menggunakan persamaan 2.1:

$$h(i) = \text{card} \{ (u, v) \vee I(u, v) = i \}^1 \quad (2.1)$$

$h(i)$ adalah sebuah histogram yang terdiri oleh jumlah piksel dan intensitas. karenanya $h(0)$ dapat diartikan sebagai banyaknya informasi dalam piksel sebanyak 0, $h(1)$ adalah banyaknya informasi dalam piksel sebanyak 1, dan seterusnya. Hingga mencapai $h(255)$ dimana piksel memiliki intensitas maksimal sebesar 255 atau $K - 1$. Hasil ketika melakukan penghitungan sebuah histogram adalah sebuah *vector* yang memiliki dimensi h dengan panjang K [17]. Sebuah citra dapat direpresentasikan tingkat keabuannya dalam bentuk histogram seperti pada gambar 2.4.

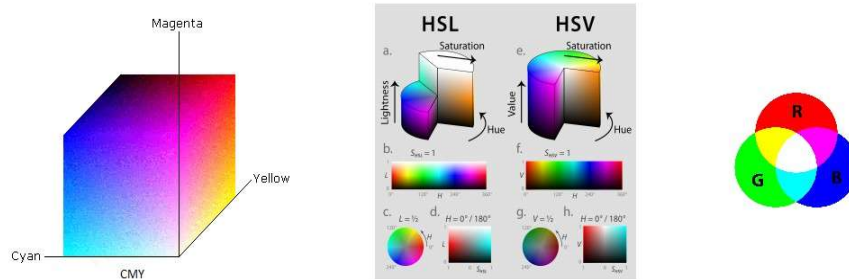


Gambar 2. 4 Representasi citra statis dalam bentuk histogram [14]

2.6 MODEL WARNA

Colorspace adalah cara agar warna dapat diciptakan, divisualisasikan dan dikhususkan dengan aturan tertentu, mata manusia dapat mencirikan sebuah warna berdasarkan atribut – atribut visual yang ditangkap oleh mata adalah *brightness*, *hue* dan *colorfulness*. Sedangkan komputer dapat mendeskripsikan dan membedakan warna menggunakan jumlah dari merah, hijau dan biru emisi fosfor yang diperlukan untuk mencocokkan sebuah warna. Sebuah hasil print dapat menghasilkan sebuah warna tertentu dalam sebagai hasil pantulan dan penyerapan dari tinta *cyan*, *magenta*, kuning dan hitam [18].

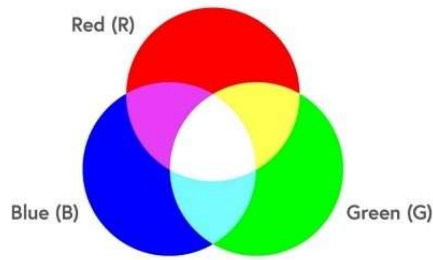
Color space atau yang lebih sering disebut model warna adalah spesifikasi sistem koordinat dan suatu subruang dalam sistem tersebut dengan setiap warna dinyatakan dengan satu titik di dalamnya. Tujuan dibentuknya ruang warna adalah untuk memfasilitasi spesifikasi warna dengan standar yang sama. Ruang warna yang paling dikenal pada perangkat komputer adalah *RGB*, yang sesuai dengan watak manusia dalam menangkap warna. Seiring perkembangan zaman, kemudian dibuat banyak ruang warna antara lain *HSI*, *CMY*, *LUV*, dan *YIQ* yang memiliki fungsi masing – masing [18].



Gambar 2. 5 *Color Space* [20]

2.7 MODEL WARNA RED GREEN BLUE (RGB)

Model warna *RGB* adalah sebuah color space dengan 3 warna primer seperti merah (*Red*), hijau (*Green*), dan biru (*Blue*), yang telah dikombinasikan untuk menghasilkan bermacam – macam warna turunan atau biasa disebut sebagai warna sekunder. Warna yang dihasilkan dengan mengkombinasikan tiga warna primer dan akan menghasilkan 8 bit merah, 8 bit hijau, dan 8 bit biru [19]. Sehingga dengan menggunakan 3 warna primer dapat menghasilkan spektrum warna yang luas mulai dari hitam jika menggabungkan 3 warna dengan nilai 0 hingga warna putih dengan mengkombinasikan 3 warna primer dengan nilai maksimal (255), model warna *RGB* dapat digambarkan seperti pada gambar 2.6.



Gambar 2. 6 Model Warna RGB [20]

Pada model warna *RGB* terdapat dua jenis warna yaitu warna primer dan warna sekunder, warna sekunder merupakan warna campuran dari 3 warna seperti warna merah, biru, kuning, hijau, magenta, dan cyan. Tabel 2.2 menunjukkan nilai kombinasi untuk menghasilkan sebuah warna.

Tabel 2. 2 Rentang Intensitas Warna pada Model Warna RGB

Warna	Rentang Warna		
	<i>Red</i>	<i>Green</i>	<i>Blue</i>
Kuning	75 – 225	102 – 255	0 – 50
Magenta	75 – 255	0 – 230	128 – 255
Biru	0 – 240	0 – 248	112 – 255
Cyan	0 – 224	128 – 255	20 – 255
Merah	128 – 255	0 – 160	0 – 128
Hijau	0 – 173	100 – 255	0 – 170

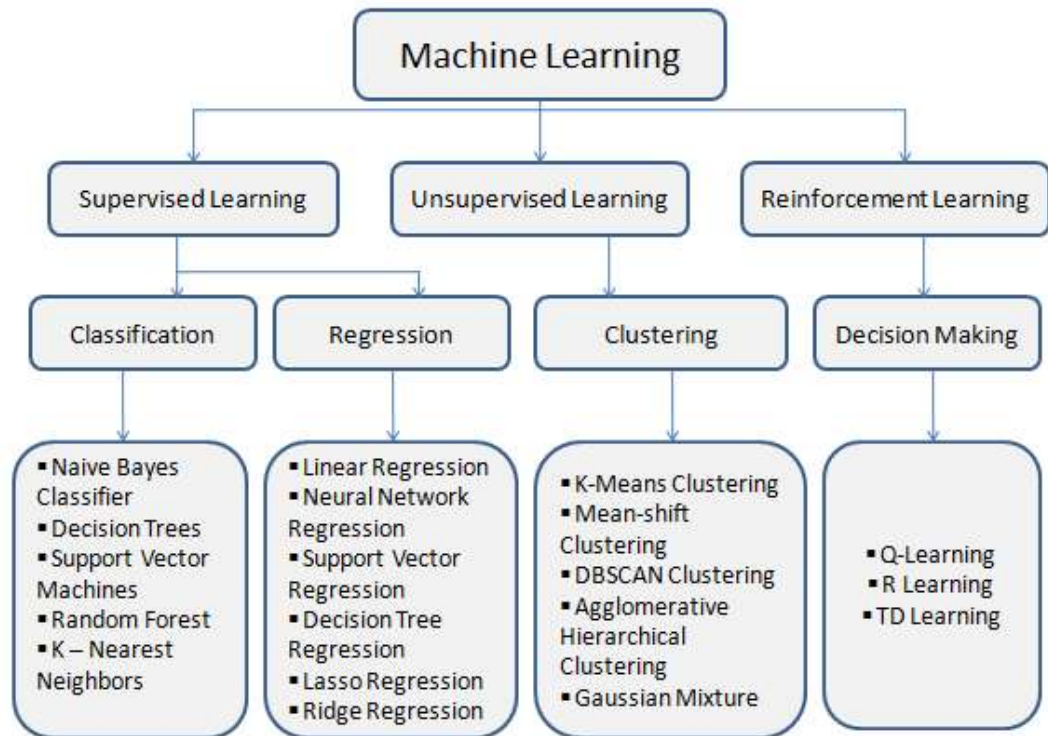
2.8 MACHINE LEARNING

Machine learning adalah salah satu pengembangan dari disiplin ilmu kecerdasan buatan, yang menggunakan pemrograman untuk menghasilkan sistem yang cerdas serta dapat melakukan pembelajaran secara mandiri, dan dapat meningkatkan pemahamannya melalui pengalaman secara otomatis [21].

Machine learning adalah studi yang berfokus untuk membuat sebuah sistem yang mampu melakukan pembelajaran secara mandiri hingga dapat melakukan pemecahan masalah tanpa harus diprogram ulang setiap kali ingin memutuskan sesuatu. Dengan membuat sistem yang mampu belajar sendiri menjadikan mesin mampu mengambil keputusan serta dapat melakukan adaptasi ketika terjadi perubahan dalam kasus tertentu. *Machine learning* dapat bekerja dengan baik apabila tersedia data sebagai input untuk dilakukan analisis terhadap kumpulan data yang lebih besar (*big data*) dan dilakukan perbandingan atau pengukuran hingga menemukan pola tertentu. Data merupakan bahan *input* krusial yang akan digunakan untuk melakukan pembelajaran (*training*) sehingga mesin dapat melakukan analisis yang benar. Didalam *machine learning* dikenal *data training* dan *data testing*, *data training* adalah data yang digunakan untuk melatih algoritma dalam *machine learning* untuk melakukan sesuatu sedangkan

data testing adalah data yang digunakan untuk mengetahui performa dari algoritma dalam *machine learning* yang telah dilatih ketika menemukan data baru yang belum pernah diberikan dalam *data training* [21].

Machine learning menggunakan teknik untuk menangani data besar (*big data*) dengan cara yang cerdas untuk memberikan hasil yang tepat. Berdasarkan teknik pembelajarannya, *machine learning* dapat dibedakan menjadi 3 tipe pembelajaran yaitu *supervised learning*, *unsupervised learning*, dan *reinforcement learning*. *Supervised learning* adalah salah satu pembelajaran dalam *machine learning* yang memanfaatkan data training yang sudah diberi label (*labeled data*) dalam proses pelatihan dan pembelajaran, sehingga mesin mampu mengidentifikasi data baru yang sudah diberi label maupun data yang belum diberi label untuk kemudian melakukan klasifikasi atau prediksi. Jenis – jenis *machine learning* dapat dilihat pada gambar 2.7.

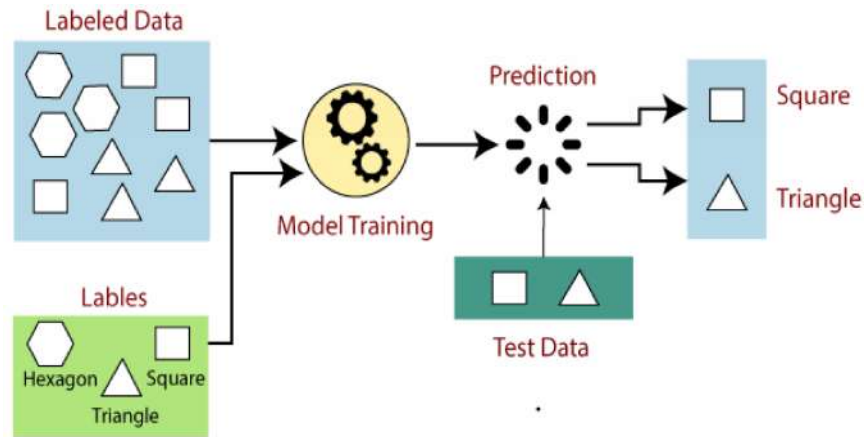


Gambar 2. 7 Jenis – jenis *Machine Learning* [22]

2.9 SUPERVISED LEARNING

Supervised learning adalah salah satu metode dalam *machine learning* yang memanfaatkan data *training* yang sudah berlabel (*labeled data*) guna melakukan pelatihan pada mesin [21], metode *supervised learning* didasarkan pada data *training* yaitu kumpulan sampel data yang memiliki label dan kriteria tertentu yang berbeda dari data label lain, data – data sampel tersebut digunakan untuk

menyimpulkan karakteristik distribusi dan ukuran perilaku dalam setiap jenis data yang memiliki label hingga membentuk sebuah model perilaku dari data sehingga mesin mampu mengidentifikasi label *input* dengan menggunakan fitur yang dimiliki untuk selanjutnya melakukan prediksi dari data *input* maupun klasifikasi data dari data yang ada di dalam *dataset training*. Diagram blok *Supervised learning* dapat dilihat pada gambar 2.8.



Gambar 2. 8 Supervised Machine Learning [23]

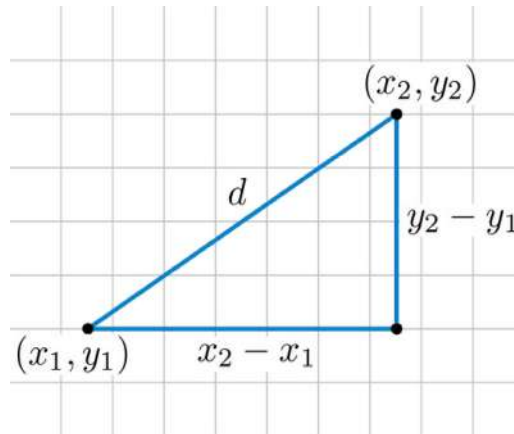
2.10 K- NEAREST NEIGHBOR

Algoritma *K-Nearest Neighbor* adalah salah satu cara yang digunakan untuk mengklasifikasikan sebuah objek dengan cara mengelompokkan data baru berdasarkan jarak data tersebut dengan data yang memiliki kemiripan dengannya atau biasa disebut dengan data tetangga (*neighbor*). Data data yang ada diproyeksikan kedalam ruang dimensi banyak, yang berisikan nilai – nilai masing masing data. Untuk kemudian dilakukan pencarian terhadap data tetangganya untuk melakukan klasifikasi, pada K-NN nilai k terbaik dalam suatu kasus tergantung pada jenis data dan jumlah data. Namun pada umumnya, nilai dari K yang lebih besar mampu mengurangi efek *noise* pada klasifikasi, namun akan membuat batasan antar setiap data klasifikasi menjadi lebih kabur [8]. Dalam melakukan klasifikasi perlu dilakukan pengukuran jarak dari sebuah data ke data yang lain untuk menentukan tingkat kemiripannya dan mengetahui tetangga (*neighbor*) dari data tersebut, dalam mengukur jarak terdapat beberapa cara untuk mengukur jarak kedekatan antar data baru dengan data lama maupun dengan data uji cara tersebut diantaranya adalah *Chebyshev Distance*, *Euclidean distance*, dan *Manhattan Distance*, sistem pengukuran jarak yang digunakan dalam penelitian ini adalah *Euclidean distance*.

2.11 EUCLIDEAN DISTANCE

Euclidean distance adalah salah satu metode pengukuran jarak guna mengukur tingkat kemiripan citra satu terhadap citra lain. Metode ini

menggunakan representasi citra di dimensi banyak (*hyperspace*) secara numerik hingga dapat melakukan pengukuran untuk mengukur jarak kemiripan dua citra dengan melakukan operasi matematika. Metode *Euclidean distance* dapat digambarkan menggunakan vektor 2 dimensi yang merepresentasikan nilai dari 2 citra yang berbeda. Vektor ini kemudian direpresentasikan dalam diagram kartesian x dan y. Dimana masing-masing dari vektor tersebut memiliki pada nilai pada bidang x dan y [24].



Gambar 2. 9 Euclidean distance [24]

Pada Gambar 2.9 dapat disimpulkan bahwa untuk mencari jarak dari kedua vektor tersebut dilakukan dengan menarik garis diagonal yang lurus dari 1 titik ke titik yang lainnya. Penggambaran ini mirip dengan cara mencari resultan dari dua buah vektor. Sehingga jarak dapat dituliskan menjadi rumus:

$$E_d = \sqrt{\sum_1^x \sum_1^y i \cdot i} \quad (2.2)$$

Besar jarak pengukuran antara citra A ke citra B dapat dihitung apabila melakukan pengurangan nilai masing – masing piksel pada citra A ke citra B di koordinat *pixel* yang sama secara berulang ulang dari awal *pixel* sampai akhir *pixel*, berarti metode ini hanya dapat dilakukan apabila kedua citra memiliki nilai *pixel* yang sama. Dalam kasus membandingkan citra satu ke citra lainnya metode penghitungan jarak *Euclidean distance* cocok digunakan untuk melakukan pengolahan citra [24].

2.12 CONFUSION MATRIX

Confusion matrix adalah salah satu cara pengukuran untuk mengukur tingkat keandalan pada sebuah sistem klasifikasi dengan melihat hasil sistem [25]. *Confusion matrix* dapat ditunjukkan seperti Tabel 2.3 berikut:

Tabel 2.3 Confusion Matrix

N=200		Nilai Lapangan	
		<i>True</i>	<i>False</i>
Nilai Prediksi	<i>True</i>	<i>TP</i>	<i>FP</i>
	<i>False</i>	<i>FN</i>	<i>TN</i>

Keterangan:

- TP (*True Positive*) adalah jumlah data cabai matang yang kemudian terbaca matang oleh sistem ketika dilakukan pengujian.
- FP (*False Positive*) adalah jumlah data cabai mentah namun terbaca oleh sistem sebagai matang ketika dilakukan pengujian.
- TN (*True Negative*) adalah jumlah data cabai belum matang yang kemudian terbaca belum matang oleh sistem ketika dilakukan pengujian.
- FN (*False Negative*) adalah banyaknya data cabai matang namun kemudian terbaca belum matang oleh sistem ketika dilakukan pengujian.

2.13 AKURASI

Akurasi adalah pengujian yang melakukan perbandingan antara nilai prediksi dan nilai lapangan dengan cara menghitung total data yang secara benar dilakukan klasifikasi oleh sistem dengan nilai total percobaan [25]. Persamaan akurasi ditunjukkan oleh persamaan berikut:

$$Accuracy = \left(\frac{TP + TN}{TP + TN + FP + FN} \right) \times 100\% \quad (3.1)$$

2.14 PRESISI

Presisi adalah pengujian yang menggunakan data yang secara benar diklasifikasikan dibandingkan dengan total jumlah data yang dianggap benar oleh sistem [25]. Persamaan presisi ditunjukkan oleh persamaan berikut:

$$Precision = \left(\frac{TP}{TP + FP} \right) \times 100\% \quad (3.2)$$

2.15 RECALL

Recall adalah pengujian yang membandingkan nilai data yang secara benar diklasifikasikan dengan total data klasifikasi benar ditambahkan dengan data kesalahan sistem [25]. Persamaan recall ditunjukkan oleh persamaan berikut:

$$Recall = \left(\frac{TP}{TP + FN} \right) \times 100\% \text{ (3.3)}$$