

## BAB II DASAR TEORI

### 2.1 TINJAUAN PUSTAKA

Penelitian dengan judul “*Klasifikasi Jenis Bawang Menggunakan Metode K-Nearest Neighbor Berdasarkan Ekstraksi Fitur Bentuk dan Tekstur*” yang ditulis oleh J. A. Widians, H. S. Pakpahan, E. Budiman, Haviluddin dan M. Soleha. Penelitian tersebut bertujuan untuk klasifikasi jenis bawang berdasarkan ekstraksi fitur daun. Metode yang digunakan adalah ekstraksi fitur Histogram dibandingkan dengan metode ekstraksi fitur dengan klasifikasi *K-Nearest Neighbor* (KNN). Hasil penelitian menunjukkan bahwa dengan Berdasarkan 5 kali pengujian dengan menggunakan k3, k5, k7 hasil terbaik yaitu pada pengujian 50% dengan jumlah 25 data *training* dan 25 data *testing* dari 50 citra bawang dengan akurasi sebesar 84%. Penggunaan nilai k terbaik pada pengujian ke 3 yaitu 70% dengan k7 mendapatkan akurasi sebesar 86.66%, dan rata-rata akurasi nilai k pada 5 kali pengujian nilai akurasi yang tertinggi yaitu pada k7 dengan akurasi sebesar 83.13% [3].

Penelitian Hasan Bisri, M. Arief Bustomi, Endah Purwanti pada tahun 2013 yang berjudul “*Klasifikasi Citra Paru-Paru dengan Ekstraksi Fitur Histogram dan Jaringan Syaraf Tiruan Backpropagation*” meneliti tentang pengklasifikasikan citra paru – paru dengan ekstrasi fitur histogram menggunakan jaringan syaraf tiruan *backpropagation*. Penelitian pada sistem jaringan syaraf tiruan ini adalah dengan mengubah nilai parameter - parameternya. Parameter-parameter tersebut adalah jumlah neuron pada lapisan tersembunyi (*hidden layer*) dan nilai *epoch* jaringan. Nilai-nilai parameter-parameter inilah yang diubah untuk mendapatkan hasil terbaik. Hasil pelatihan sistem ini berubah-ubah, karena nilai bobot awal untuk bobot awal input, bobot awal bias input, bobot awal lapisan dan bobot awal bias lapisan diatur untuk terisi secara acak, sehingga tingkat akurasi yang didapat juga berubah. Dari hasil pengujian untuk beberapa kali perubahan parameter, didapatkan hasil rata-rata yang terbaik adalah sebesar 98,89 %. Dari hasil pelatihan dengan tingkat akurasi yang terbaik ini, kemudian bobot dari jaringan yang diperoleh disimpan dan dijadikan parameter untuk pengujian sistem terhadap citra yang belum pernah dilatihkan pada jaringan. Dengan nilai parameter yang sama yaitu

$epoch = 370$ , nilai  $error = 0,001$ ,  $learning\ rate = 0,2$  serta jumlah *neuron* sebanyak 2500, jaringan diuji dengan menggunakan data citra uji. Hasil pengujian ini menghasilkan tingkat akurasi rata-rata sebesar 65% [6].

Penelitian dengan judul “Klasifikasi Citra Makanan menggunakan K-Nearest Neighbor dengan fitur bentuk *Simple Morphological Shape Descriptors* dan fitur warna *Grayscale Histogram*” dengan penulis M. Rizky. S menggabungkan metode Histogram untuk ekstraksi fitur dengan *K-Nearest Neighbor* untuk klasifikasi pengenalan jenis citra makanan. Penelitian Metode yang digunakan dalam melakukan klasifikasi citra menggunakan KNN dengan membandingkan nilai  $k$  yang digunakan memengaruhi nilai akurasi. Jika menggunakan nilai  $k=1$  dalam melakukan klasifikasi maka nilai akurasi yang dihasilkan sebesar 79,2%. Jika menggunakan nilai  $k=3$  dalam melakukan klasifikasi maka nilai akurasi yang dihasilkan sebesar 70,8%. Serta, jika menggunakan nilai  $k=5$  dalam melakukan klasifikasi maka nilai akurasi yang dihasilkan sebesar 54,2% [7].

Penelitian selanjutnya berjudul “Implementasi *Convolutional Neural Networks* untuk Klasifikasi Citra Tomat Menggunakan *Keras*” dengan penulis Tiara Shafira. Penelitian ini menghasilkan Implementasi metode *Convolutional Neural Network (CNN)* untuk klasifikasi citra tomat dilakukan menggunakan *package Keras* pada *software RStudio* versi 1.1.383. Banyaknya *layer* konvolusi yang digunakan yaitu sebanyak empat *layer*, fungsi aktivasi yang digunakan yaitu *ReLU*, dan beberapa penggunaan parameter lainnya [8].

Berdasarkan beberapa Tinjauan Pustaka yang dilakukan penulis menunjukkan bahwa, penggunaan metode Histogram dengan *K-Nearest neighbor* dapat digunakan untuk pembuatan model klasifikasi *Supervised Learning* (Dataset berlabel). Hasil dari penggabungan kedua metode tersebut memungkinkan menghasilkan akurasi yang baik untuk dataset *input* berupa citra tanaman atau daun.

## 2.2 DASAR TEORI

### 2.2.2 Pertanian Cerdas (*Smart Farming*)

Pertanian cerdas adalah sebuah sistem pertanian mutakhir yang didukung dengan teknologi masa kini untuk menunjang produktivitas hasil pertanian agar lebih maksimal, sistem ini bertujuan untuk mengatur dan memprediksi hasil panen serta masalah yang dihadapi oleh para petani [9]. Istilah *Precision farming* atau *smart farming* (pertanian cerdas) mulai dikenal. Beberapa peneliti telah mendefinisikan *precision farming*, yang pada dasarnya pengertian *precision farming* adalah sistem manajemen pertanian yang bertujuan untuk meningkatkan produktivitas dan penggunaan sumberdaya baik melalui peningkatan hasil atau berkurangnya input dan efek lingkungan yang merugikan dengan memanfaatkan teknologi informasi [10].

Jenis Teknologi Pertanian Cerdas (*Smart farming tech / SFT*) dibagi menjadi tiga kategori utama yang, seperti disebutkan di atas, mencakup sistem siklus pertanian presisi:

1. Teknologi akuisisi data: kategori ini berisi semua teknologi survei, pemetaan, navigasi, dan penginderaan.
2. Teknologi analisis dan evaluasi data: teknologi ini berkisar dari model keputusan sederhana berbasis komputer hingga manajemen pertanian yang kompleks dan sistem informasi termasuk banyak variabel berbeda.
3. Teknologi aplikasi presisi: kategori ini berisi semua teknologi aplikasi, dengan fokus pada aplikasi tingkat variabel dan teknologi panduan

Teknologi akuisisi data pada sistem pertanian cerdas digunakan untuk merekam dan memetakan bidang dan karakteristik tanaman dibagi ke dalam kategori di bawah ini:

1. Teknologi sistem navigasi satelit (sebenarnya teknologi ini merekam posisi sebenarnya yang dapat digunakan untuk tujuan yang berbeda seperti panduan, pemetaan, dll.)
2. Teknologi pemetaan
3. Akuisisi data properti lingkungan (Pencitraan berbasis kamera, sensor kelembapan tanah)
4. Mesin dan propertinya teknologi sistem navigasi satelit [10].

### **2.2.3 Bawang Merah**

Bawang merah merupakan komoditas umbi yang sangat penting di Indonesia. Komoditas ini sering digunakan sebagai bumbu penyedap dalam masakan maupun sebagai obat karena khasiatnya. Luas panen bawang merah di Indonesia pada tahun 2016 yaitu 149.635 ha (meningkat 22,53 % dari tahun 2015) dengan luas areal panen terbesar terletak di Jawa Tengah yaitu 53 ha. Sebaliknya produktivitas bawang merah secara umum menurun sekitar 3,89 % dibandingkan tahun 2015 (BPS, 2018). Peningkatan permintaan bawang merah terus meningkat sesuai dengan pertumbuhan jumlah penduduk [11].

Bentuk batang bawang merah seperti cakram, beruas dan diantara ruas terdapat kuncup-kuncup. Batang bagian bawah merupakan tempat tumbuhnya akar dan bagian atas merupakan umbi semu. Daun bawang merah bertangkai pendek, berbentuk bulat mirip pipa, berlubang, panjang lebih dari 45 cm, meruncing pada bagian ujung dan bawahnya melebar seperti kelopak dan membengkak. Daun bawang merah berwarna hijau muda bergantung varietas dan saat siap panen daun berubah menguning layu dan akhirnya mengering dimulai dari bagian bawah tanaman. Bunga bawang merah terdiri atas tangkai dan tandan bunga. Setiap tangkai terdapat lebih dari 50-200 kuntum bunga, setiap bunga memiliki 5-6 benang sari dan putik dengan daun bunga yang berwarna hijau bergaris keputih-putihan atau putih dan bakal buah. Bawang merah juga memiliki biji yang masih muda berwarna putih dan setelah tua menjadi hitam dan berbentuk pipih. Bawang merah memiliki konsentrasi nitrogen dalam jaringan sebesar 1,25% [12].

Bawang merah dipanen setelah umurnya cukup tua, yaitu pada umur 60 hari. Tanaman bawang merah dipanen setelah terlihat tanda-tanda seperti daun telah rebah atau leher batang lunak, tanaman rebah dan daun menguning. Pemanenan dilaksanakan pada saat tanah kering dan cuaca cerah untuk menghindari adanya serangan penyakit busuk umbi pada saat umbi disimpan [13].

### **2.2.4 Citra Digital**

Citra digital merupakan gambar dua dimensi yang dihasilkan dari analog dua dimensi yang kontinu menjadi gambar melalui proses sampling. Gambar analog dibagi menjadi N baris dan M kolom sehingga menjadi gambar diskrit. Citra digital merupakan citra yang dapat diolah komputer [14]. Yang disimpan dalam komputer

hanyalah angka-angka yang menunjukkan besar intensitas pada masing-masing piksel. Karena berbentuk data numerik, maka citra digital dapat diolah dengan komputer tetapi tidak dengan citra yang bukan dari data numerik maka tidak dapat diolah dengan komputer.

Citra digital dapat dibedakan menjadi tiga yaitu citra biner (*binary image*), citra keabuan (*grayscale image*) dan citra warna (*color image*).

### 1. *Color Image* atau RGB (*Red, Green, Blue*)

Citra berwarna (*color image*) atau bisa disebut citra RGB adalah jenis citra yang menyajikan warna dalam bentuk komponen merah, hijau, dan biru. Setiap komponen warna menggunakan 8 bit (nilainya berkisar antara 0 sampai dengan 255). Dengan demikian, kemungkinan warna yang bisa disajikan mencapai 2553 atau 16.581.375 (16K) warna [15].

### 2. *Grayscale* (Keabu-abuan)

Piksel dari citra digital *grayscale* (keabu-abuan) memiliki gradasi warna mulai dari putih sampai hitam. Rentang warna pada *grayscale image* banyak digunakan dalam dunia kedokteran (X-ray). *Grayscale* merupakan hasil rata-rata dari *color image* dengan persamaan sebagai berikut:

$$I_{BW}(x, y) = \frac{I_R(x, y) + I_G(x, y) + I_B(x, y)}{3} \quad (2.1)$$

Keterangan:

$I_R(x, y)$  : nilai piksel *Red* titik (x,y)

$I_G(x, y)$  : nilai piksel *Green* titik (x,y)

$I_B(x, y)$  : nilai piksel *Blue* titik (x,y)

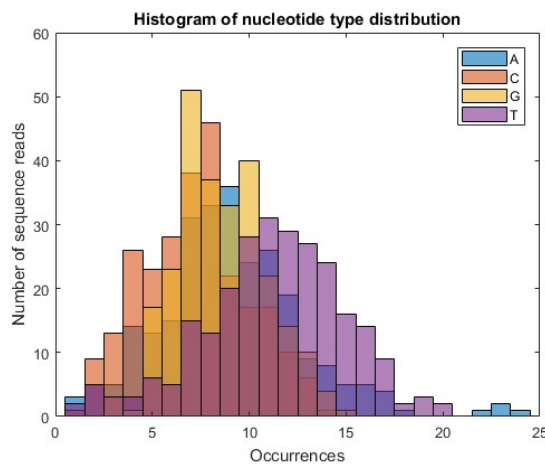
$I_{BW}(x, y)$  : nilai piksel *Black and White* titik (x,y)

### 3. *Binary Image*

Komponen warna dari *binary image* hanya terdiri dari dua warna yaitu hitam dan putih, maka dibutuhkan 1 bit per piksel (0 dan 1) atau apabila dalam 8 bit (0 dan 255). *Binary image* cocok digunakan pada penerapan teks, sidik jari, maupun gambar arsitektur [16].

### 2.2.5 Histogram

Histogram adalah sebuah metode yang digunakan untuk menggambarkan atau mempresentasikan statistik gambar dalam bentuk visual agar mudah ditafsirkan dalam bentuk format yang dibutuhkan. Dengan histogram dapat dimanfaatkan dan dioptimalkan untuk menentukan jenis masalah tertentu dalam sebuah gambar. Pada kasus ini histogram digunakan untuk membantu menentukan kesiapan panen dari suatu tanaman, membantu meningkatkan visual penampilan gambar dan sebagai media untuk menganalisis perubahan dari daun suatu tanaman[17].



Gambar 2.1 Statistik gambar dalam bentuk visual histogram [18].

Histogram menunjukkan distribusi piksel berdasarkan intensitas *graylevel* (derajat keabuan) yang dimiliki oleh tiap-tiap piksel. Penggunaan histogram sebagai metode ekstraksi ciri didasarkan pada perbedaan sebaran atau distribusi piksel yang terjadi di antara *frame* stoma, *frame* sebagian stoma, dan *frame* bukan stoma. Pada metode ekstraksi ciri histogram, *bin* merupakan banyaknya batang warna yang akan terbentuk, atau menunjukkan jumlah pembagian rentang warna pada histogram. Jumlah titik ekstraksi ciri yang dihasilkan oleh suatu histogram adalah sama dengan jumlah *bin* yang digunakan pada histogram tersebut[19].

Secara umum, histogram di definisikan sebagai sebuah frekuensi distribusi piksel – piksel yang mendeskripsikan suatu intensitas atau kemungkinan spesifik pada suatu citra. Konsep histogram adalah frekuensi dari sebuah intensitas, artinya citra dengan kanal lebih dari 1 dapat memiliki histogram untuk tiap – tiap kanalnya, sedangkan untuk citra 1 kanal (*grayscale*) memiliki 1 buah histogram. Sebuah

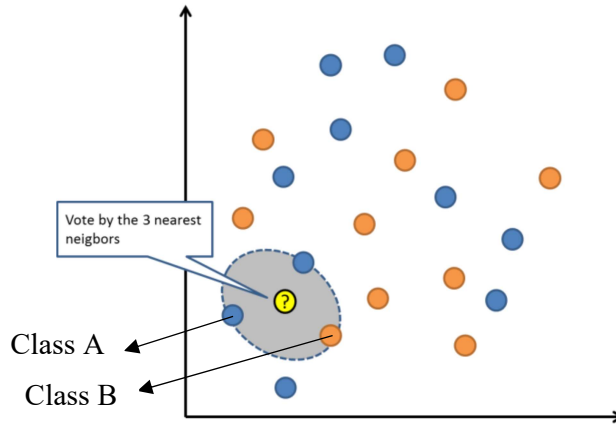
histogram ( $h$ ) untuk sebuah citra *grayscale* mempunyai nilai intensitas dengan rentang  $I_{(u,v)} \in [0, K - 1]$  dimana dimana nilai dari  $K$  adalah intensitas dari derajat keabuan citra yaitu  $K = 2^8 = 256$ . Sebuah histogram didefinisikan dengan persamaan sebagai berikut [20]:

$$h(i) = \text{card} \{(u, v) | I(u, v) = i\}^1 \quad (2.2)$$

dimana  $h(i)$  merupakan histogram yang tersusun oleh banyaknya piksel dan intensitas. Oleh karena itu,  $h(0)$  adalah banyaknya piksel dengan nilai 0,  $h(1)$  banyaknya piksel dengan nilai 1, dan seterusnya. Hingga akhirnya  $h(255)$  adalah semua piksel putih dengan intensitas maksimum 255 atau sama dengan  $K - 1$ . Hasil dari komputasi histogram adalah vektor satu dimensi  $h$  dengan panjang  $K$  [21].

### 2.2.6 *K-Nearest Neighbor (K-NN)*

*K-Nearest Neighbor (KNN)* mulai ditemukan dan digunakan sejak awal tahun 1970-an untuk pemahaman pola dan perkiraan statistik sebagai teknik *non-parametrik*, sekarang ini mulai banyak digunakan untuk klasifikasi objek maupun prediksi objek [22].



Gambar 2.2 Klasifikasi *K-Nearest Neighbor* dalam bentuk visual [23].

Algoritma KNN merupakan salah satu algoritma klasifikasi yang paling terkenal digunakan untuk memprediksi kelas dari catatan atau (sampel) dengan kelas yang tidak ditetapkan berdasarkan kelas dari catatan neighbor atau tetangganya. algoritma ini terbuat dari tiga langkah sebagai berikut [24].

- a) Menghitung jarak *record* masukan dari semua catatan pelatihan.
- b) Mengatur catatan pelatihan berdasarkan jarak dan pemilihan K-tetangga terdekat.
- c) Menggunakan kelas yang memiliki mayoritas diantara k-tetangga terdekat (metode ini menganggap kelas sebagai kelas record input yang diamati lebih dari semua kelas-kelas lain antar K-tetangga terdekat) [24].

Algoritma *K-Nearest Neighbor* adalah salah satu metode klasifikasi pengambilan keputusan berdasarkan jarak terdekat pada data nilai dengan persamaan yang digunakan matriks satuan jarak atau biasanya menggunakan *Euclidean*.

Data diproyeksikan keruang dimensi banyak, dimana masing-masing dimensi merepresentasikan fitur dari data. Ruang inii dibag menjadi bagian-bagian berdasarkan klasifikasi data. Nilai *K* yang terbaik untuk algoritma ini tergantung pada data. Secara umum, nilai dari *K* yang tinggi akan mengurangi efek *noise* pada klasifikasi, tetapi membuat Batasan antarsetiap klasifikasi menjadi lebih kabur. Ada beberapa cara untuk mengukur jarak kedekatan antar data baru dengan data lama diantaranya *Euclidean distance*, *Manhattan Distance*, *Chebysev Distance*, yang sering digunakan adalah *Euclidean distance*. [25].

### **2.2.7 Model Warna RGB**

Model warna RGB adalah contoh warna yg terdiri atas tiga warna utama yaitu *Red* (Merah), *Green* (Hijau), *Blue* (Biru), yg sudah ditambahkan memakai banyak sekali cara untuk membuat bermacam – macam warna. Warna yg didapatkan dari kombinasi antara 3 warna tadi & masing – masing mempunyai nilai 8 bit merah, 8 bit hijau, & 8 bit biru. Campuran warna utama menggunakan porposisi seimbang akan membuat perbedaan makna warna kelabu, jika 3 warna tadi disaturasikan penuh maka akan membuat warna putih [26].























Tabel 2.1 Intensitas warna RGB Warna primer dan Sekunder[26]

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

Pada model warna RGB terdapat dua warna primer dan warna sekunder yang ditandai dengan adanya warna merah, hijau, biru, kuning, magenta dan cyan. Tabel 2.1 menunjukkan kisaran intensitas warna dari setiap warna. Warna hijau dengan spektrum hijau dapat dilihat pada pita hijau dari saluran hijau yang ditunjukkan pada Tabel 2.2.

Tabel 2.2 Spektrum Rentang Warna Hijau untuk Seluruh Kanal [27]

<i>Color</i>	<i>HTML / CSS Color Name</i>	<i>Hex Code #RRGGBB</i>	<i>Decimal Code (R,G,B)</i>
	<i>Lawgreen</i>	#7CFC0D	rgb(124, 252, 13)
	<i>Chartreuse</i>	#7FFF0D	rgb(127, 255, 13)
	<i>Limegreen</i>	#32CD32	rgb(50, 205, 50)
	<i>Lime</i>	#00FF00	rgb(0, 255, 0)
	<i>Forestgreen</i>	#228B22	rgb(34, 139, 34)
	<i>Green</i>	#008000	rgb(0, 128, 0)
	<i>Darkgreen</i>	#006400	rgb(0, 100, 0)
	<i>Greenyellow</i>	#ADFF2F	rgb(173, 255, 47)
	<i>Yellowgreen</i>	#9ACD32	rgb(154, 205, 50)
	<i>Springgreen</i>	#00FF7F	rgb(0, 255, 127)
	<i>Mediumspringgreen</i>	#00FA9A	rgb(0, 250, 154)
	<i>Lightgreen</i>	#90EE90	rgb(144, 238, 144)
	<i>Palegreen</i>	#98FB98	rgb(152, 251, 152)
	<i>Darkseagreen</i>	#8FBC8F	rgb(143, 188, 143)
	<i>Mediumseagreen</i>	#3CB371	rgb(60, 179, 113)
	<i>Lightseagreen</i>	#20B2AA	rgb(32, 178, 170)
	<i>Seagreen</i>	#2E8B57	rgb(46, 139, 87)
	<i>Olive</i>	#808000	rgb(128, 128, 0)
	<i>Darkolivegreen</i>	#556B2F	rgb(85, 107, 47)
	<i>Olivdrab</i>	#6B8E23	rgb(107, 142, 35)

### 2.2.8 Euclidean Distance

Jarak *Euclidean* adalah perhitungan yang digunakan untuk mengukur jarak dua titik dalam ruang *Euclidean* yang mendalami tentang hubungan antara sudut

dan jarak Dalam matematika, jarak *Euclidean* digunakan sebagai metode untuk mengukur dua titik dalam satu dimensi, sehingga memberikan hasil yang sama dengan perhitungan Pitagoras .

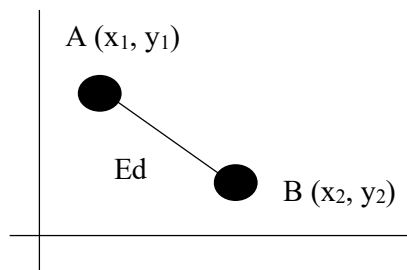
Berikut adalah persamaan *euclidean distance* :

$$E_d = \sqrt{\sum_1^x \sum_1^y (Ax, y - Bx, y)^2} \quad (2.3)$$

Keterangan :

$d$  = Jarak  
 $x$  = Koordinat latitude  
 $y$  = Koordinat Longitude

Jarak *Euclidean* adalah metode klasifikasi objek. *Euclidean distance* merupakan salah satu metode jarak yang digunakan untuk mengukur kemiripan satu citra dengan citra lain jika berada dekat dengan citra tersebut. Metode ini diperoleh dengan merepresentasikan citra dalam beberapa dimensi (*hyperspace*) sehingga jarak antara dua citra yang berbeda dapat diatur. Metode jarak *Euclidean* dapat dipahami langkah demi langkah dengan menggambarkan vektor 2 dimensi yang mewakili 2 gambar yang berbeda. Kedua vektor ini direpresentasikan dalam plot x dan y *Cartesian*. Masing masing dari vektor tersebut memiliki pada nilai pada bidang x dan y.



Gambar 2.3 Penggambaran metode *Euclidean distance* [28]

Gambar 2.1 menunjukkan dengan jelas menunjukkan bahwa jarak dua vektor diperoleh dengan menggambar garis diagonal dari satu titik ke titik lainnya. Representasi ini mirip dengan mencari hasil dari dua vektor. [28]

### 2.2.9 K-Fold Cross Validation

K-Fold cross validation adalah metode yang digunakan untuk membagi data menjadi data latih dan data uji. Cross-Classifier pada K-fold membagi sampel data secara acak menjadi K subset independen. Subset digunakan sebagai data uji dan subset K1 sebagai data latih. Proses klasifikasi silang akan diulang hingga K kali. Data asli dibagi secara acak menjadi K subset independen yaitu  $S_1, S_2, \dots, S_k$ , ukuran setiap subset adalah sama, Metode ini menjadi solusi untuk memastikan bahwa setiap data mendapatkan kesempatan yang sama untuk menjadi data uji dan data latih.

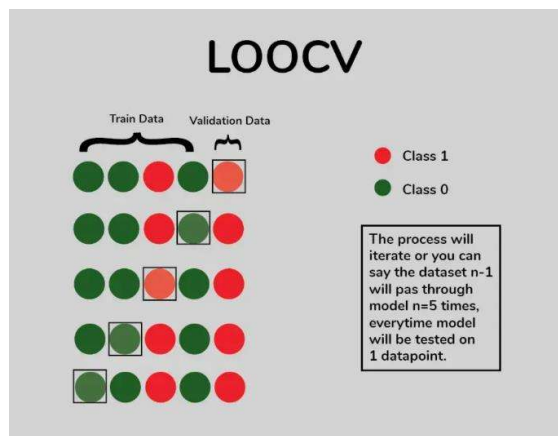
## K-Fold Cross-Validation

Percobaan 1	Test	Train	Train	Train	Train
Percobaan 2	Train	Test	Train	Train	Train
Percobaan 3	Train	Train	Test	Train	Train
Percobaan 4	Train	Train	Train	Test	Train
Percobaan 5	Train	Train	Train	Train	Test

Gambar 2.4 Visualisasi *K-Fold Cross Validation* [29].

### 2.2.10 Leave One Out Cross Validation (LOOCV)

Bentuk umum dari algoritma ini disebut validasi *K-Fold Cross*, yang membagi kumpulan data menjadi k kumpulan data dengan ukuran yang sama. Setiap kali dijalankan, satu *shard* bertindak sebagai set data validasi sementara *shard* lainnya menjadi set data pelatihan.



Gambar 2.5 Visualisasi *Leave One Out Cross Validation (LOOCV)* [29].

Prosedur dieksekusi  $k$  kali sehingga setiap level data berpotensi menjadi level data validasi tepat satu kali dan level data latih  $k-1$  kali. Bentuk khusus dari metode ini adalah ketika  $k$  didefinisikan  $k = N$ , jumlah data dalam kumpulan data. Metode ini disebut *breakoneout*, yaitu *dataset* pengujian hanya berisi sebagian data, sedangkan pembelajaran (dan validasi) dilakukan sebanyak  $N$  kali.