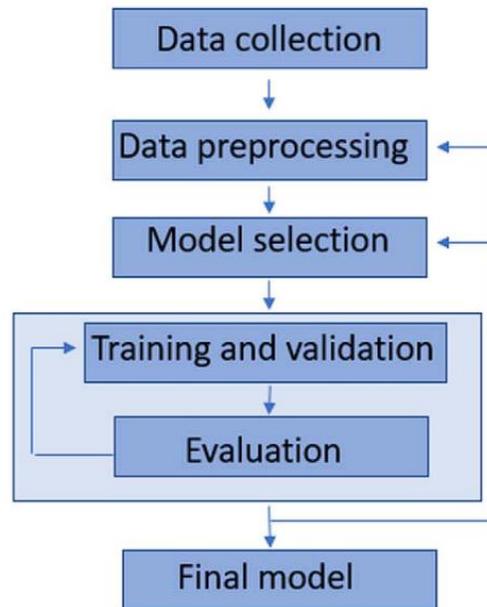


## BAB II TINJAUAN PUSTAKA

### 2.1 Machine Learning

Machine learning adalah salah satu cabang kecerdasan buatan yang berkaitan dengan pengembangan dan studi algoritma statistik yang dirancang untuk dapat belajar secara mandiri dari data, kemudian menggunakan pembelajaran tersebut untuk menangani data baru yang belum pernah ditemui sebelumnya, sehingga sistem ML dapat menjalankan tugas-tugas tertentu tanpa memerlukan instruksi eksplisit yang terprogram sebelumnya. Alur kerja machine learning bisa dilihat pada gambar 2.1.



Gambar 2. 1 Alur kerja Machine Learning

Sumber : (Ng et al., 2020)

#### 1. Data collection

Data collection atau pengumpulan data adalah proses pengumpulan sebuah data sebuah objek yang akan diteliti untuk memperoleh informasi mengenai subjek tertentu untuk tujuan pengambilan keputusan.

## 2. Data preprocessing

Data preprocessing adalah proses menghasilkan data mentah untuk model pembelajaran mesin. Prapemrosesan data sering kali memiliki dampak signifikan pada kinerja generalisasi algoritma ML yang diawasi (Kotsiantis & Kanellopoulos, 2006). Data mentah umumnya terdapat kesalahan, nilai yang hilang, dan mungkin dalam format yang tidak dapat digunakan yang tidak dapat langsung digunakan untuk model pembelajaran mesin. Data preprocessing merupakan tugas yang diperlukan untuk membersihkan data dan membuatnya sesuai untuk model pembelajaran mesin yang juga meningkatkan akurasi dan efisiensi model pembelajaran mesin. Pada penelitian ini data preprocessing hanya digunakan untuk mengubah ukuran pixel dataset dan mengubah format dataset menjadi .png semua.

## 3. Model selection

Model selection atau Pemilihan model adalah proses pembelajaran mesin yang berfokus pada penggunaan algoritma dan model yang paling sesuai untuk set data tertentu. Proses ini mencakup penilaian dan perbandingan model yang berbeda untuk mengidentifikasi model yang memberikan hasil terbaik. Berbagai fitur dipertimbangkan, dan berbagai metrik digunakan untuk mencapai kesimpulan.

## 4. Training and validation

Training data adalah kumpulan data aktual yang digunakan untuk melatih model, dimana model melihat dan belajar dari data ini untuk memprediksi hasil atau membuat keputusan yang tepat. Sebagian besar data pelatihan dikumpulkan dari beberapa sumber dan kemudian diproses dan diorganisir untuk memberikan kinerja model yang tepat. Jenis data pelatihan sangat menentukan kemampuan model untuk menggeneralisasi, yaitu semakin baik kualitas dan keragaman data pelatihan, semakin baik pula kinerja model. Data ini lebih dari 60% dari total data yang tersedia untuk proyek.

Validation digunakan untuk menyempurnakan hiperparameter model dan sebagai bagian dari pelatihan model. Model hanya melihat data ini untuk evaluasi tetapi tidak belajar dari data ini, sehingga memberikan evaluasi objektif yang tidak bias terhadap model. Dataset validasi dapat digunakan untuk regresi juga dengan mengganggu pelatihan model ketika kehilangan dataset validasi menjadi lebih besar daripada kehilangan dataset pelatihan, yaitu mengurangi bias dan varians. Data ini sekitar 10-15% dari total data yang tersedia untuk proyek, tetapi ini dapat berubah tergantung pada jumlah hiperparameter, yaitu jika model memiliki cukup banyak hiperparameter, maka menggunakan set validasi yang besar akan memberikan hasil yang lebih baik. Sekarang, setiap kali akurasi model pada data validasi lebih besar dari pada data pelatihan maka model dikatakan telah menggeneralisasi dengan baik.

## 5. Evaluation

Evaluasi model adalah proses yang menggunakan beberapa metrik yang membantu kita untuk menganalisis kinerja model. Seperti yang kita semua tahu bahwa pengembangan model adalah proses multi-langkah dan pemeriksaan harus dilakukan untuk mengetahui seberapa baik model menggeneralisasi prediksi di masa depan. Oleh karena itu, evaluasi model memainkan peran penting agar kita dapat menilai kinerja model kita. Evaluasi juga membantu untuk menganalisis kelemahan utama model. Ada banyak metrik seperti Akurasi, Presisi, Recall, skor F1, Area di bawah Kurva, Matriks Kebingungan, dan Mean Square Error. Pada penelitian ini menggunakan evaluasi mode Akurasi, Presisi, Recall, dan Skor F1.

Untuk Rumus yang digunakan masing model evaluasi sebagai berikut:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\textit{Precision} = \frac{TP}{TP + FP}$$

$$\textit{Recall} = \frac{TP}{TP + FN}$$

$$\textit{F1 - Score} = 2 \times \frac{\textit{Precision} \times \textit{Recall}}{\textit{Precision} + \textit{Recall}}$$

TP = True Positif

TF = True False

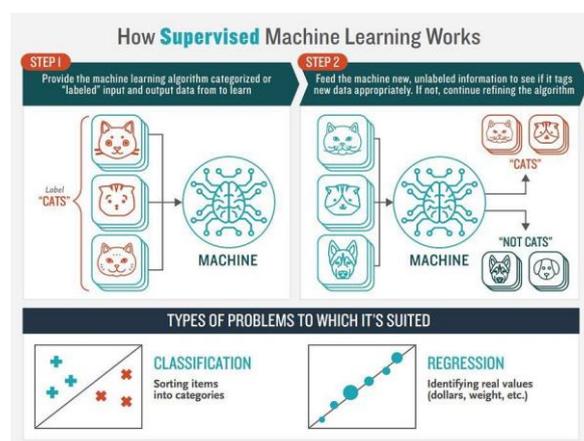
FP = False Positif

FN = False Negatif

Adapun beberapa tipe machine learning:

### 2.1.1 Supervised Learning

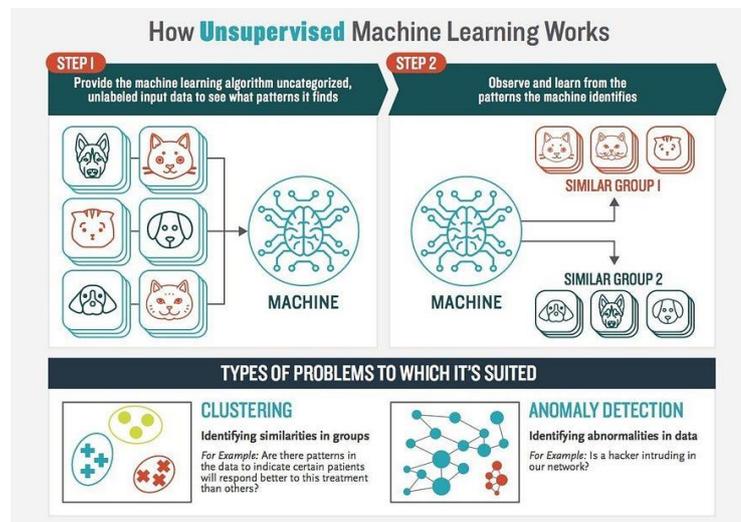
Supervised learning merupakan sistem pembelajaran yang di mana sebuah sistem akan diberikan input dan akan dilatih dengan berbagai data untuk dapat menghasilkan suatu output (Mitaart, 2020). Tujuan dari supervised learning adalah sebuah sistem yang dapat memprediksi suatu data yang diberikan berdasarkan hasil pembelajaran yang telah dilakukan.



Gambar 2. 2 Gambaran Cara kerja Supervised Learning

### 2.1.2 Unsupervised Learning

Unsupervised learning yaitu sebuah sistem yang tidak diajarkan untuk dapat memproses suatu input menjadi output. Cara pengklasifikasian unsupervised learning sendiri menggunakan kemiripan atribut yang ada dari sebuah data (Mitaart, 2020).



Gambar 2. 3 Gambaran Cara kerja Unsupervised Learning

## 2.2 Klasifikasi

Klasifikasi adalah proses menemukan model (atau fungsi) yang mendeskripsikan dan membedakan kelas atau konsep data (Han et al., 2011). Klasifikasi sendiri melalui dua proses yaitu proses pembelajaran dan proses kedua ialah proses klasifikasi.

Klasifikasi gambar sendiri adalah mendeskripsikan keseluruhan gambar dengan mengekstraksi fitur secara manual atau menggunakan *Machine Learning* yang kemudian menggunakan pengklasifikasi untuk mengidentifikasi kategori objek (Nawrocka et al., 2021).

### 2.3 Buah Jeruk

Tanaman jeruk adalah tanaman buah tahunan yang berasal dari Asia. Cina dipercaya sebagai tempat pertama kali jeruk tumbuh. Jeruk memiliki berbagai kandungan gizi yang sangat tinggi, tidak hanya vitamin C dalam 180 gr buah jeruk, juga terdapat nutrisi lainnya seperti : Protein, Kalori dan Serat yang sangat tinggi. Jeruk merupakan buah yang sering dikonsumsi oleh kebanyakan masyarakat Indonesia, dimana yang dilansir dari Kementerian Koordinator Bidang Perekonomian Republik Indonesia pada tahun 2021, dimana produksi buah jeruk siam mencapai 2,4 juta ton pada tahun tersebut (Limanseto, 2022).



Gambar 2. 4 Buah Jeruk Manis

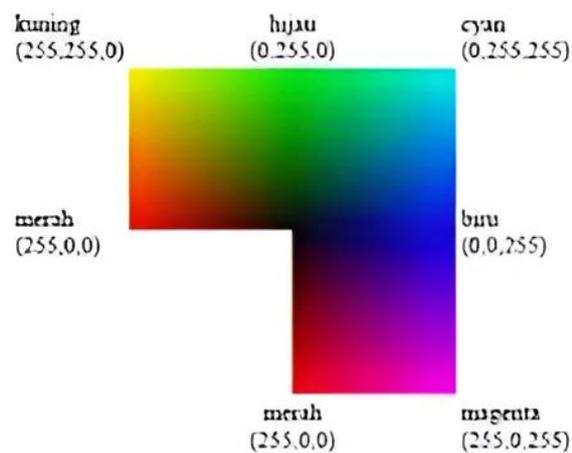
Sumber : ([www.astronauts.id](http://www.astronauts.id))

### 2.4 Citra

Citra adalah representasi visual dua dimensi yang terbentuk dari citra analog dua dimensi yang kontinu melalui proses sampling yang mengubahnya menjadi citra diskrit. Proses sampling dibagi menjadi dua yaitu downsampling dan upsampling. Downsampling yaitu mengurangi jumlah piksel atau resolusi spasial citra yang nantinya menghasilkan citra dengan nilai yang lebih rendah. Sementara itu, upsampling adalah meningkatkan jumlah piksel atau meningkatkan resolusi citra (Yelliy N, 2019).

## 2.5 Citra berwarna

Citra warna atau yang sering dikenal sebagai citra RGB terdiri dari tiga lapisan yang mewakili setiap piksel, yaitu R (Red), G (Green), dan B (Blue) (Setyawan & Mendrofa, 2021). Kombinasi dari intensitas warna merah, hijau, dan biru menentukan warna dari setiap piksel. Setiap lapisan menggunakan warna delapan bit dengan nilai berkisar antara 0 hingga 255, sehingga format file grafis akan menyimpan citra warna ini sebagai 24 bit. Jenis warna ini mampu menampilkan grafik kualitas tinggi dengan 16.581.375 warna.



Gambar 2. 5 Representasi warna RGB pada citra

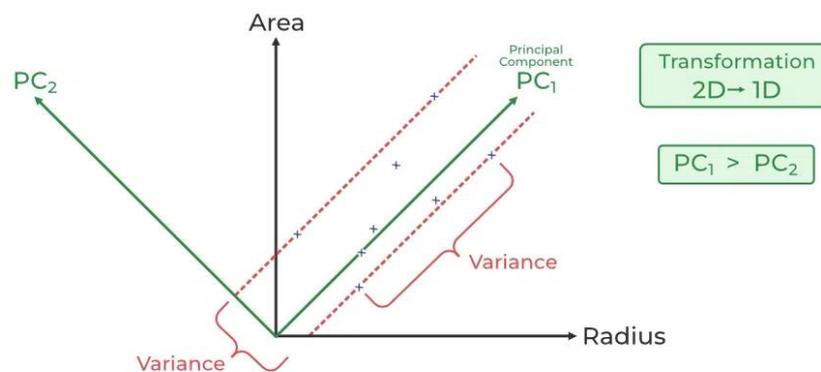
## 2.6 Pengolahan citra digital

Pengolahan citra adalah suatu metode yang digunakan untuk memproses gambar dalam bentuk 2 dimensi. Pengolahan citra juga bisa melakukan operasi untuk memperbaiki, menganalisa, atau mengubah suatu gambar (Setyawan & Mendrofa, 2021). Tujuan dari pengolahan citra sendiri yaitu menganalisis suatu gambar sehingga didapatkan informasi baru tentang gambar tersebut. Adapun tujuan pengolahan citra digital yaitu :Memperbaiki kualitas gambar dilihat dari aspek radiometrik (peningkatan kontras<sup>1</sup>, transformasi warna, restorasi citra) dan dari aspek geometrik (rotasi, translasi, skala, transformasi geometrik)<sup>2</sup>, Melakukan proses

penarikan informasi atau deskripsi objek atau pengenalan objek yang terkandung pada citra<sup>3</sup>, Melakukan kompresi atau reduksi data untuk tujuan penyimpanan data<sup>4</sup>, transmisi data, dan waktu proses data<sup>5</sup>.

## 2.7 PCA

Principal component analysis (PCA) adalah suatu teknik statistik multivariat yang secara linear mengubah bentuk sekumpulan variabel asli menjadi kumpulan variabel yang lebih kecil. PCA digunakan untuk mengurangi dimensi kumpulan data dengan menemukan kumpulan variabel baru yang lebih kecil dari kumpulan variabel awal dan tetap mempertahankan sebagian besar informasi sampel yang nantinya berguna untuk regresi dan klasifikasi data.



Gambar 2. 6 Principal Component Analysis

Sumber : (<https://www.geeksforgeeks.org>)

PCA menganalisis table data yang mewakili pengamatan dijelaskan oleh beberapa variabel dependen, yang secara umum saling berkorelasi. Tujuannya adalah untuk mengekstrak informasi penting dari data tabel dan untuk mengekspresikan informasi ini sebagai satu set baru variabel ortogonal yang disebut komponen utama. Komponen utamanya adalah kombinasi linier dari variabel asli dalam kumpulan data dan diurutkan berdasarkan tingkat kepentingannya. Total varians yang ditangkap oleh

semua komponen utama sama dengan total varians pada dataset asli. Komponen utama (Principal Component) pertama menangkap variasi terbanyak dalam data, namun komponen utama kedua menangkap varian maksimum yang ortogonal terhadap komponen utama pertama, dan seterusnya seperti yang bisa dilihat pada gambar 2.2.

Adapun algoritma yang diterapkan pada setiap proses pada PCA sebagai berikut (Akbar et al., 2022):

a. Membentuk matriks parameter citra

Hal pertama yang harus dilakukan dalam penerapan algoritma PCA yaitu menentukan matriks parameter citra yang akan dilatih. Seperti yang terlihat pada persamaan dibawah ini :

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1n} \\ x_{21} & x_{22} & x_{23} & & x_{2n} \\ x_{31} & x_{32} & x_{33} & & x_{3n} \\ \dots & & & & \dots \\ x_{m1} & x_{m2} & x_{m3} & \dots & x_{mn} \end{bmatrix}$$

Citra berukuran  $m \times n$  dimana  $x$  adalah data,  $m$  adalah jumlah dari data citra training dan  $n$  adalah jumlah citra tes.

b. Mencari rata – rata seluruh citra

Setelah proses pembentukan matriks selesai maka selanjutnya dilakukan pencarian rata – rata dari citra, dimana tujuan dari dilakukannya pencarian nilai rata – rata sendiri yaitu untuk mengetahui *noise* yang dapat mengurangi keakuratan dalam penghitungan PCA nantinya. Rumus yang bisa dipakai sebagai berikut:

$$\tau = \frac{(x_{12}+x_{21}+x_{31}+\dots+x_{m1})}{m} \dots\dots\dots(1)$$

$$= \frac{\sum_{j=1}^m x_{ji}}{m} = [\tau_1, \tau_2, \dots, \dots, \dots, \tau_n] \dots\dots\dots(2)$$

c. Penggandaan rata – rata

Proses ini bertujuan untuk menyamakan dimensi nilai rata – rata dengan dimensi nilai pada citra sampel. Persamaan yang bisa dipakai sebagai berikut:

$$\mu = \alpha x \tau \dots\dots\dots(3)$$

Dimana  $\alpha$  adalah matriks kolom dengan dimensi  $m \times 1$ .

d. Perhitungan nilai rata – rata nol

Tujuan dari perhitungan nilai rata – rata nol sendiri yaitu untuk mengurangi noise yang mengganggu pada saat dilakukan perhitungan pada PCA nantinya. Persamaan yang bisa diterapkan sendiri yaitu:

$$\phi = \chi - \mu \dots\dots\dots(4)$$

e. Pembentukan matriks kovarian

Setelah didapatkan nilai rata – rata nol tadi selanjutnya dilakukan pembentukan matriks kovarian. Adapun rumus persamaan yang bisa diterapkan sebagai berikut:

$$c = \frac{1}{m-1} \phi_{ji} * \phi^T j i \dots\dots\dots(5)$$

f. Menentukan matriks *eigenvalue* dan *eigenvector*

Langkah selanjutnya setelah matriks kovarian diperoleh adalah menentukan matriks *eigenvalue* dan *eigenvector*. Rumus persamaan yang bisa dipakai yaitu sebagai berikut:

$$C - Z = |C - \lambda I| \dots\dots\dots(6)$$

$$|C - \lambda I| = 0 \dots\dots\dots(7)$$

Dimana C = matriks kovarian.

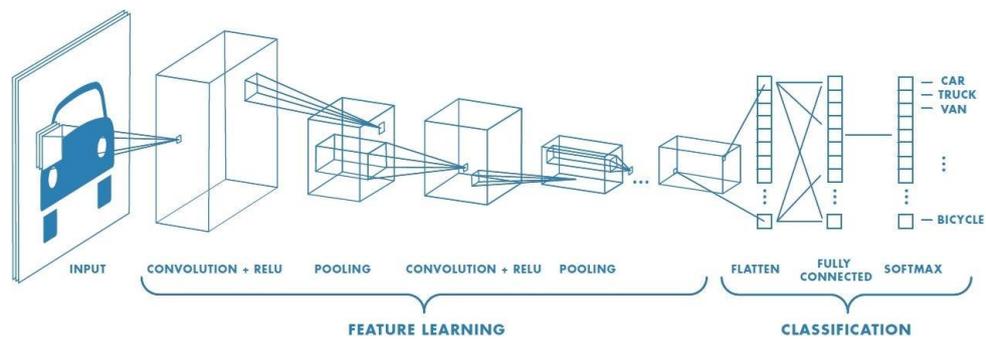
Z = matriks *eigenvalue*,

$\lambda = \text{Scalar}$  pembentukannya

$I =$  matriks identitasnya

## 2.8 CNN

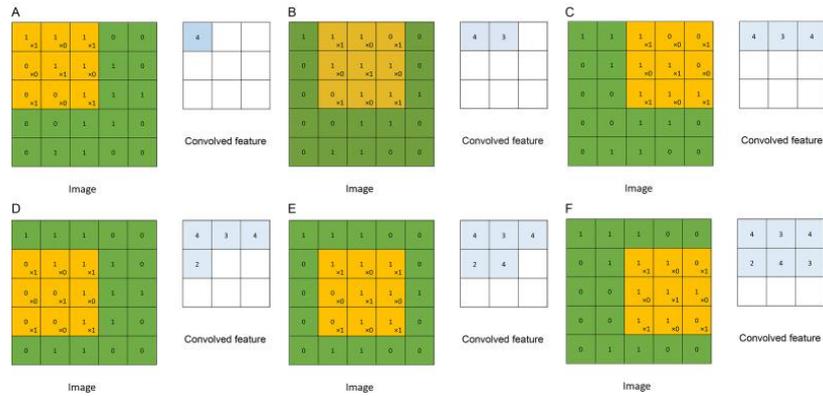
Algoritma CNN merupakan pengembangan dari metode sebelumnya yaitu Multiplayer Perceptron (MLP) yang didesain untuk tujuan mengolah data dua dimensi. Algoritma ini bekerja dengan mengidentifikasi hal-hal yang sifatnya visual, misalnya saja mengenali objek pada gambar. Sistem ini terdiri dari tiga bagian, yakni Convolutional Layer, Pooling Layer, dan Fully-Connected (FC) layer dengan tugas masing-masing, mulai dari mengidentifikasi hal yang ringan sampai dengan kompleks.



Gambar 2. 7 Sistematika algoritma CNN

### 2.5.1 Convolutional Layer

Convolutional Layer sendiri adalah bagian dari tahapan arsitektur CNN itu sendiri. Pada tahapan ini dilakukan operasi konvolusi output dari layer sebelumnya. Operasi Konvolusi sendiri adalah operasi yang dilakukan pada dua fungsi argumen bernilai nyata dan operasi ini mengadaptasi fungsi output sebagai feature map dari inputan citra. Input dan Output dinilai sebagai dua argumen bernilai riil. Pada konteks citra, operasi konvolusi bekerja dengan cara melakukan pergerakan pada kernel ke berbagai posisi sesuai dengan ukuran citra. Cara kerjanya seperti pada gambar 2.8 berikut



Gambar 2. 8 Convolutional Operation

Sumber : (ScientDirect.com)

**2.5.2 Stride**

Stride merupakan suatu parameter yang menentukan besaran dari pergeseran filter pada proses convolutional filter. Seperti pada gambar 2.8 dimana besaran stride yaitu 3x3 maka convolutional filter akan bergeser sebanyak 3 pixel secara vertikal maupun horizontal. Semakin Kecil stride yang digunakan maka detail informasi yang didapatkan dari inputan semakin banyak, tapi memerlukan waktu yang cukup lama juga (Qolbiyatul Lina, 2019).

**2.5.3 Padding**

Padding adalah suatu parameter yang menentukan jumlah pixel (berisi nilai 0) dimana nantinya akan ditambahkan pada setiap sisi dari inputan. Tujuannya sendiri yaitu untuk memanipulasi dimensi output dari convolutional layer agar tetap sama dengan dimensi input. Nantinya akan meningkatkan performa dari model convolutional layer dimana nantinya akan fokus pada informasi yang sebenarnya yang berada diantara zero padding. Untuk menghitung dimensi dari convolutional layer bisa menggunakan rumus sebagai berikut:

$$Output = \frac{W+N+2P}{s} + 1 \dots\dots\dots(8)$$

Keterangan :

W : Panjang/Tinggi Input

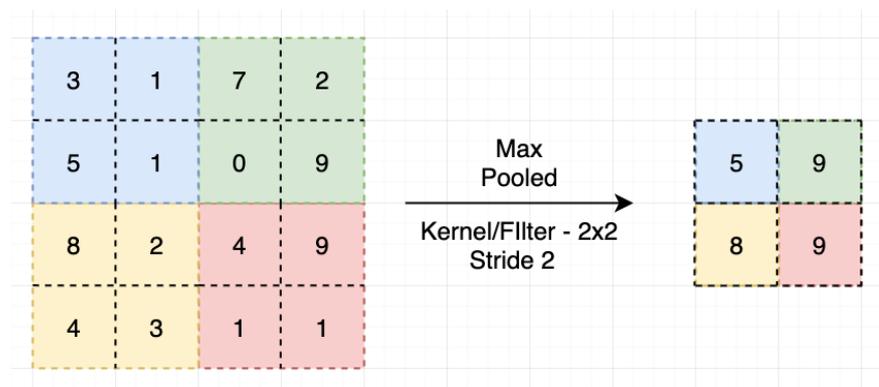
N : Panjang/Tinggi Filter

P : Zero Padding

S : Stride

### 2.5.4 Pooling Layer

Pooling layer biasanya berada setelah convolutional layer. Pooling layer sendiri terdiri dari sebuah filter dengan ukuran dan stride tertentu yang bergeser pada seluruh area feature map. Tujuan dari penggunaan pooling layer yaitu untuk mengurangi dimensi dari feature map (downsampling), sehingga mempercepat komputasi karena parameter yang harus di update semakin sedikit dan mengatasi overfitting (Qolbiyatul Lina, 2019). Pooling yang paling sering digunakan adalah Max Pooling dan Average Pooling.

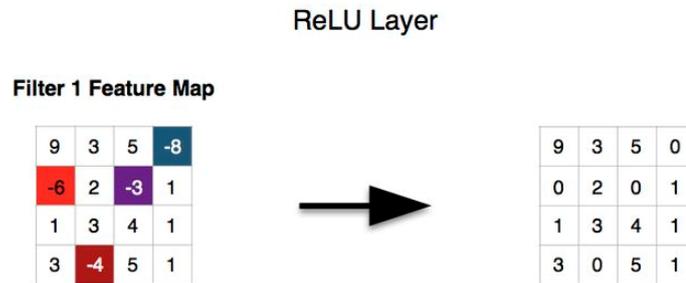


Gambar 2. 9 Operasi Max Pooling

### 2.5.5 Relu Layer

ReLU atau Rectified Activation Function merupakan non-saturating activation function yang dapat meningkatkan non-linearitas decision function dan network secara keseluruhan, tanpa harus mempengaruhi bidang-bidang reseptif pada convolution layer (Muhammad Yunus, 2020). ReLU sangat efektif untuk menghapus

nilai negative pada feature/activation map dan dijadikan 0. Untuk gambaran cara kerja Reu Layer bisa dilihat pada gambar 2.10



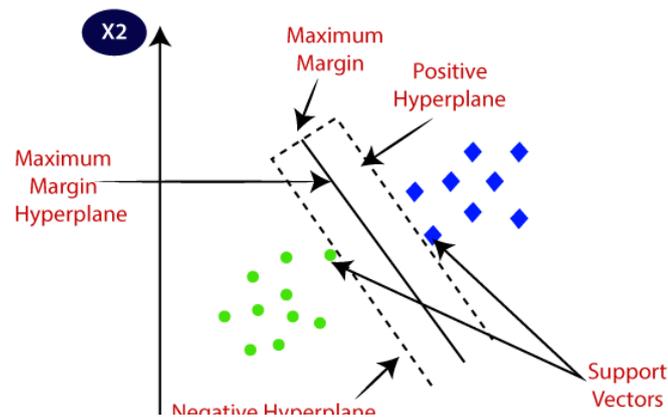
Gambar 2. 10 Cara Kerja Relu Layer

### 2.5.6 Fully Connected

Fully-connected adalah lapisan dimana semua neuron aktivitas dari lapisan sebelumnya terhubung semua dengan neuron di lapisan selanjutnya. Setiap aktivitas dari lapisan sebelumnya perlu diubah menjadi data satu dimensi sebelum dapat dihubungkan ke semua neuron di lapisan Fully-Connected.

## 2.9 SVM

Support Vector Machine (SVM) adalah algoritma *Machine Learning* yang diawasi dan dapat digunakan untuk klasifikasi dan regresi. Cara kerja SVM didasarkan pada SRM atau Structural Risk Minimization yang dirancang untuk mengolah data menjadi Hyperplane yang mengklasifikasikan ruang input menjadi dua kelas. Cara kerja algoritma SVM diawali dengan



Gambar 2. 11 Model SVM

pengelompokan kasus-kasus linier yang dapat dipisahkan dengan hyperplane dan dibagi menurut kelasnya.

Pada gambar 2.11 merupakan model sederhana dari algoritma SVM, dimana model ini terdapat dua pola yang berbeda dan dua pola tersebut dipisahkan oleh SVM. Dimana kedua pola tersebut dipisahkan sesuai dengan nilai datanya, jika datanya bernilai negative maka akan berada pada negative hyperlane dan jika datanya bernilai positif maka akan berada pada positive hyperlane hal tersebut bisa terjadi secara sempurna oleh hyperlane berdimensi  $d$  (Nugroho et al., 2003). Adapun persamaan yang tunjukan untuk menentukan hyperlane :

$$\bar{w} \cdot \bar{x} + b = 0 \dots\dots\dots(9)$$

keterangan:

$w$  : *weight* / nilai dari bidang normal

$x$  : data input

$b$  : bias / posisi bidang relatif pada pusat koordinat

Pattern  $x_i$  yang termasuk pada class data yang bernilai negatif dapat dirumuskan sebagai berikut:

$$\bar{w} \cdot \bar{x}_i + b \leq -1 \dots\dots\dots(10)$$

Sedangkan pattern  $x_i$  yang termasuk pada class data positif dirumuskan sebagai berikut:

$$\bar{w} \cdot \bar{x}_i + b \geq +1 \dots\dots\dots(11)$$

Dan Jarak titik data  $x$  ke hyperplane dapat dihitung sebagai proyeksi  $x$  ke vektor normal  $w$ , dibagi besaran  $w$

$$d = \frac{|w^T x + b|}{|w|} \dots\dots\dots(12)$$

Dimana :

$w$  : *weight* / nilai dari bidang normal

$x$  : data input

$b$  : bias / posisi bidang relatif pada pusat koordinat

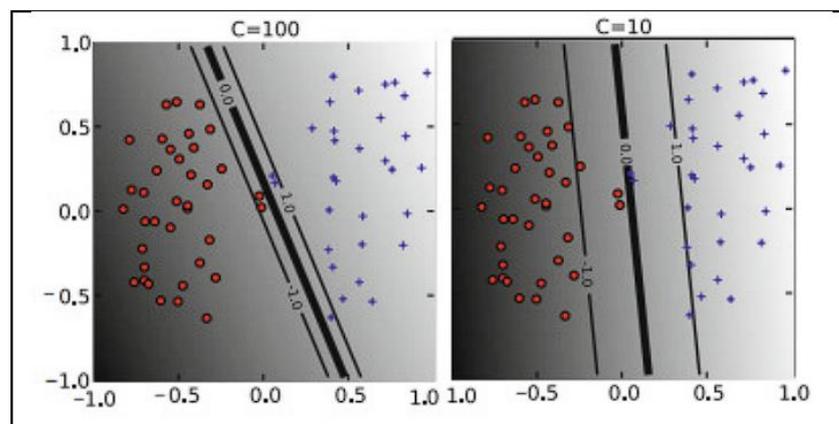
$d$  = jarak  $x$  ke hyperlane

Hasil akurasi dari sebuah model dihasilkan pada tahap transisi menggunakan SVM sangat dipengaruhi oleh jenis fungsi kernel dan parameter yang digunakan (Praghakusma & Charibaldi, 2021). Ada beberapa fungsi kernel yang dipakai pada sebuah aplikasi untuk mengatasi masalah pada Support Vector Machine (SVM) (Nanda et al., 2018):

a. Linier kernel

Persamaan yang bisa dipakai yaitu:

$$K(x_i, x) = x_i x$$



Gambar 2. 12 Kernel Linier

Pada gambar 2.12 penggunaan nilai  $C$  dengan nilai 10 akan menghasilkan nilai margin error yang rendah serta akan memperlebar nilai margin dan mengabaikan ignore point terdekat dengan decision

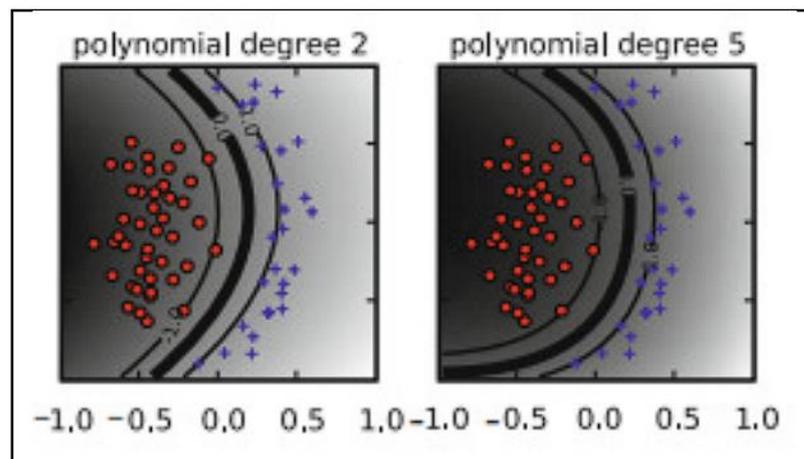
boundary, begitu sebaliknya jika nilai C yang tinggi (100) akan menghasilkan margin error besar dan mempersempit margin.

b. Polynomial Kernel

Persamaan yang bisa dipakai yaitu:

$$K(x_i, x) = (y(x_i^T x) + r)^p \dots \dots \dots (13)$$

Dimana  $x_i$  adalah data latih,  $x$  adalah data uji,  $p$  adalah derajat polinomial.



Gambar 2. 13 Kernel Polynomial

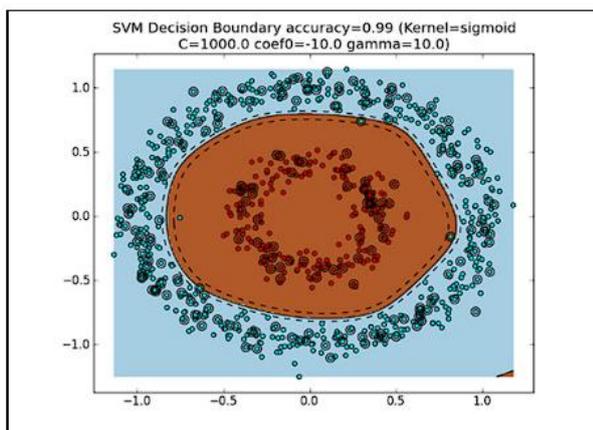
Derajat atau degree pada kernel polinomial mengontrol fleksibilitas dari hasil klasifikasi. Penggunaan derajat degree yang terlalu tinggi pada kernel polinomial dapat menyebabkan overfitting atau model yang digunakan terlalu fokus pada data latih, sehingga pada pengujian dengan data uji yang berbeda akan menyebabkan penurunan akurasi (Kowalczyk, 2017).

c. Sigmoid Kernel

Persamaan yang bisa dipakai yaitu:

$$K(x_i, x) = \tanh(y(x_i^T x) + r) \dots \dots \dots (14)$$

Dimana  $x_i$  merupakan data latih,  $x$  adalah data uji dan  $r$  adalah koefisien.



Gambar 2. 14 Kernel Sigmoid

Penggunaan gamma yang terlalu tinggi pada kernel sigmoid dapat menurunkan tingkat akurasi suatu klasifikasi tetapi tetap bergantung pada jumlah fitur yang digunakan, semakin banyak jumlah fitur maka gamma yang digunakan cenderung kecil begitupun sebaliknya (Al-Mejibli et al., 2020).

## 2.10 Matlab

MATLAB adalah sebuah bahasa pemrograman tingkat tinggi yang secara khusus digunakan untuk komputasi numerik, pemrograman, dan visualisasi. MATLAB memungkinkan manipulasi matriks, pem-plot-an fungsi dan data, implementasi algoritme, pembuatan antarmuka pengguna, dan peng-antarmuka-an dengan program dalam bahasa lainnya. Meskipun hanya bernuansa numerik, sebuah kotak kakas (toolbox) yang menggunakan mesin simbolik MuPAD, memungkinkan akses terhadap kemampuan aljabar komputer.

## 2.11 Confusion Matrix

Confusion matrix adalah alat yang digunakan untuk menggambarkan kinerja model klasifikasi pada data uji yang sudah diketahui hasil sebenarnya. Tujuan utama dari confusion matrix yaitu untuk memvisualisasikan dan menganalisis hasil prediksi yang dibuat oleh model,

sehingga dapat lebih mudah memahami kelebihan dan kekurangan model dalam mengklasifikasikan data (Rina, 2023).

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Gambar 2. 15 Confusion Matrix Table

Pada gambar 2.15 terdapat empat 4 istilah untuk merepresentasikan hasil proses klasifikasi pada confusion matrix. Keempat istilah tersebut adalah True Positive (TP), True Negative (TN), False Positive (FP) dan False Negative (FN). Dimana TP merupakan data positif yang diprediksi benar, TN merupakan data negatif yang diprediksi benar, FP merupakan kondisi dimana data negatif atau data yang salah diprediksi positif atau benar, dan FN merupakan kondisi dimana data tersebut benar namun diprediksi salah (Nugroho, 2019).

### 2.10.1 Akurasi

Akurasi adalah metrik evaluasi yang mengukur seberapa baik model membuat prediksi yang benar dari total prediksi yang dilakukan. Dalam konteks klasifikasi, akurasi memberikan gambaran mengenai seberapa sering model memprediksi kelas yang benar, baik itu kelas positif maupun negatif (Rina, 2023). Untuk menghitung nilai akurasi dapat menggunakan rumus matematika berikut:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

### 2.10.2 Presisi

Presisi adalah tingkat keakuratan antara informasi yang diminta oleh user dan respon yang diberikan oleh sistem (Hendriyana & Yazid Hilman Maulana, 2020). Dan rumus yang dipakai untuk menghitung presisi sebagai berikut :

$$\mathbf{Precision} = \frac{\mathbf{TP}}{\mathbf{TP} + \mathbf{FP}}$$

### 2.10.3 Recall

Recall adalah tingkat kesuksesan sebuah sistem dalam menemukan kembali sebuah informasi (Hendriyana & Yazid Hilman Maulana, 2020). Rumus yang dipakai untuk menghitung recall sebagai berikut:

$$\mathbf{Recall} = \frac{\mathbf{TP}}{\mathbf{TP} + \mathbf{FN}}$$

### 2.10.4 F1-Score

F1-Score adalah rata – rata dari nilai Presisi dan recall yang didapatkan (Taslim et al., 2019). Untuk rumus yang dipakai sebagai berikut:

$$\mathbf{F1 - Score} = 2 \times \frac{\mathbf{Precision} \times \mathbf{Recall}}{\mathbf{Precision} + \mathbf{Recall}}$$

## 2.12 Studi Literatur

Dalam penyusunan penelitian ini, peneliti terinspirasi dan menggunakan referensi dari beberapa penelitian sebelumnya yang terkait.

Penelitian sebelumnya (Swasono et al., 2023) menjelaskan tentang Klasifikasi Penyakit pada Citra Buah Jeruk Menggunakan Convolutional Neural Network (CNN) dengan Arsitektur Alexnet. Pada penelitian tersebut menggunakan dataset sebanyak 1790 citra buah jeruk dengan 4 kelas, setiap kelasnya terdiri dari : fresh, blackspot, canker, dan greening, dengan gambar pelatihan yang diubah menjadi 227x227 pixel. Dilakukan pemisahan data training dan data testing yaitu 90%:10%, 80%:20%, 70%:30%, setiap pemrosesan mengeksekusi 300 epoch. Didapatkan dataset dengan 90% data training dan 10% data testing memperoleh akurasi yang tinggi yaitu 94,34%.

Penelitian terkait selanjutnya (Kurniadi et al., 2021) menjelaskan tentang Analisis Perbandingan Algoritma SVM dan CNN untuk Klasifikasi Buah. Pada penelitian ini menggunakan dataset dari Kaagle.com dan didalam dataset tersebut memiliki 17 label buah dengan jumlah total data yaitu 11.219. Didapatkan hasil dari pengujian yaitu model klasifikasi SVM mendapatkan hasil akurasi sebesar 93,09% dan model klasifikasi dengan CNN mendapatkan hasil akurasi sebesar 96,87%.

Pada penelitian (Amrozi et al., 2022) mendapatkan nilai akurasi sebesar 89,86% pada penerapan algoritma SVM, dengan jumlah data dari dataset sebesar 1.256 buah, yang diklasifikasi menjadi dua jenis data yaitu pisang ambon sebanyak 656 data dan pisang lady finger sebanyak 600 data. Dengan perbandingan pembagian data 60% data latih dan 40% data uji. Dan ringkasan penelitian lainnya bisa dilihat pada tabel 2.1 berikut.

Tabel 2. 1 Studi Literatur

No	Judul Penelitian	Penulis	Sumber	Algoritma	Objek Klasifikasi	Variabel	Akurasi
1	Analisis Perbandingan Algoritma SVM dan CNN untuk Klasifikasi Buah (Kurniadi et al., 2021)	Bobby Kurniadi W, Hariyanto Prasetyo, Ghifari Ahmad L, Bagas Aditya Wibisono, Desta Sandya Prasvita	Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA) Jakarta-Indonesia (upnvj.ac.id)	SVM dan CNN	apple, avocado, banana, blueberry, cherry, corn, guava, kiwi, lemon, mango, orange, papaya, pineapple, rambutan, salak, strawberry, watermelon.	Buah, perbandingan, klasifikasi buah	SVM : 93,09% dan CNN : 96,87%
2	Klasifikasi Jenis Buah Pisang Berdasarkan Citra Warna dengan Metode SVM (Amrozi et al., 2022)	Yusuf Amrozi, Dian Yuliati, Agung Susilo, Nur Novianto, Rikza Ramadhan	Jurnal Sisfokom (Sistem Informasi dan Komputer) (atmaluhur.ac.id)	SVM	Pisang	Pisang, Citra Warna, Klasifikasi	89,86%
3	Implementasi Metode Principal Component Analysis (Pca) dan Modified K-Nearest Neighbor	Nurdiansyah, Muliadi, Rudy Herteno, Dwi Kartini,	Jurnal Mnemonic (itn.ac.id)	PCA + M-KNN	Daun Tanaman Herbal	Daun Tanaman herbal, citra daun, akurasi	89%

	Pada Klasifikasi Citra Daun Tanaman Herbal (Nurdiansyah et al., 2024)	Irwan Budiman					
4	Pengenalan Pola Batik Lampung Menggunakan Metode Principal Component Analysis (Septiani et al., 2022)	Mega Septiani, Zaenal Abidin, Permata	Jurnal Informatika dan Rekayasa Perangkat Lunak (JATIKA)	PCA	Batik Lampung	Citra, Pola batik lampung, analisis	-
5	Penerapan Algoritma Support Vector Machine (SVM) untuk Klasifikasi Berita Hoax Covid-19 (Ropikoh et al., 2021)	Isnin Apriyatin Ropikoh1, Rijal Abdulhakim, Ultach Enri, Nina Sulistiyowati	Journal of Applied Informatics and Computing (JAIC)	SVM	Berita Hoax Covid - 19	Hoax, text mining, klasifikasi	90,46%
6	Klasifikasi Penyakit pada Citra Buah Jeruk Menggunakan Convolutional Neural Network (CNN) dengan Arsitektur	Dwiretno Istiyadi Swasono, Mohammad Abuemas Rizq Wijaya, Muhamad	INFORMATICS JOURNAL (INFORMAL) jurnal.unej.ac.id	CNN	Penyakit Buah Jeruk	Alexnet, orange diseas, klasifikasi	94,34%

	Alexnet (Swasono et al., 2023).	Arief Hidayat					
--	---------------------------------------	------------------	--	--	--	--	--