

BAB II

LANDASAN TEORI

A. Sistem

Menurut Fithrie Soufitri (2023:33) “sistem merupakan suatu jaringan kerja dari prosedur- prosedur yang saling terkait, berkumpul dan bersama untuk melakukan sesuatu kegiatan atau untuk menyelesaikan suatu sasaran tertentu”.

Menurut Dedy Rahman Prehanto (2020:03) “Sistem merupakan bagian-bagian komponen dikumpulkan yang memiliki hubungan satu sama lain baik fisik maupun non fisik yang bersama-sama dalam berkerja demi tujuan yang dituju secara harmonis”.

Dari kedua definisi sistem di atas, dapat disimpulkan bahwa sistem adalah serangkaian prosedur yang saling terkait, baik secara fisik maupun nonfisik, yang bekerja secara harmonis untuk mencapai tujuan tertentu.

B. Informasi

Menurut Fithrie Soufitri (2023:04) “Informasi adalah sekumpulan data fakta yang diorganisasi atau diolah dengan cara tertentu sehingga mempunyai arti bagi penerima.

Menurut Dedy Rahman Prchanto (2020:12) “informasi merupakan hasil pengolahan data dengan cara tertentu sehingga lebih berarti dan berguna bagi penerimanya”.

Dari kedua pengertian di atas, dapat disimpulkan bahwa informasi adalah hasil pengolahan data atau fakta yang telah diorganisir sedemikian rupa sehingga memiliki makna dan manfaat yang signifikan bagi penerimanya.

C. Sistem Informasi

Menurut Fithrie Soufitri (2022:33) “*sistem Informasi* adalah sekumpulan komponen yang saling berhubungan, mengumpulkan atau mendapatkan, memproses, menyimpan, dan mendistribusikan informasi untuk menunjang pengambilan keputusan dan pengawasan dalam suatu organisasi”.

Menurut Indria Anggi Prameswari et al (2024:6) “*sistem informasi* adalah suatu rangkaian terintegrasi dari komponen-komponen yang saling berhubungan yang berkerja bersama-sama untuk mengumpulkan, menyimpan, mengolah, dan menyebarkan informasi guna mendukung pengambilan keputusan, koordinasi, pengendalian, analisis serta operasi dalam suatu organisasi”.

Berdasarkan pengertian di atas, dapat disimpulkan bahwa sistem informasi adalah kumpulan komponen yang terintegrasi dan saling berhubungan, yang berfungsi untuk mengelola informasi melalui proses pengumpulan, penyimpanan, pengolahan, dan distribusi. Sistem ini memiliki peran penting dalam mendukung berbagai aktivitas organisasi, seperti pengambilan keputusan, koordinasi, pengendalian, analisis, serta operasional.

D. Data

Menurut Patricia Wallace (2015: 15) "Data adalah bahan dasar bagi setiap sistem informasi, dan fakta mentah ini dapat muncul dalam berbagai bentuk dan rupa yang sangat beragam".

Dengan kata lain, data merupakan dasar dari setiap sistem informasi, yang berupa fakta mentah yang bisa muncul dalam berbagai bentuk. Data umumnya diproses melalui lima langkah utama, yaitu:

1. Input, data direkam dan disiapkan untuk diproses lebih lanjut. Proses input biasanya melibatkan data yang dimasukkan langsung oleh pengguna atau melalui medium fisik lain. Data yang dimasukkan akan diperiksa dan disunting oleh pengguna dan sistem komputer untuk memastikan akurasi.

2. Pemrosesan, data dihitung, dibandingkan, diurutkan, diklasifikasikan, dan diringkas untuk mengorganisasi, menganalisis, dan memanipulasi data, sehingga dapat diubah menjadi informasi yang bermanfaat bagi pengguna akhir. Kualitas data yang tersimpan dalam sistem informasi harus dikelola secara berkelanjutan.
3. Output, data yang sudah diproses akan diubah menjadi informasi, yang kemudian disajikan kepada pengguna akhir.
4. Penyimpanan, data diorganisasi dan disimpan dalam sistem dengan cara yang terstruktur untuk digunakan di masa depan.
5. Kontrol, memastikan semua tahap lainnya berjalan lancar, sambil memonitor dan mengevaluasi umpan balik dari setiap proses.

E. Database

Menurut Dedy Rahman Prehanto (2020:43) “*Database* adalah pengaturan data yang tujuan utamanya berupa fleksibilitas dan pengambilan data yang cepat”.

Pengguna dapat membuat basis data terkomputerisasi untuk menambahkan, memodifikasi, dan menghapus data, serta mengurutkan dan mengambil informasi dari basis data. Selain itu, pengguna juga dapat membuat formulir dan laporan berdasarkan data yang tersimpan dalam basis data.

F. Dataset

Menurut Yuliska, dan Syaliman (2020) Dataset adalah kumpulan data yang digunakan sebagai bahan untuk eksperimen penelitian.

1. Training set

Menurut Anggie Yolanda Ritonga (2024) “Training set adalah kumpulan data yang akan digunakan algoritma untuk membangun model simulasi”.

2. Validation set

Menurut Ariska Kurnia Rachmawati (2023) “Validation set adalah himpunan data yang digunakan untuk mengoptimasi saat melatih model”.

3. Testing set

Menurut Anggie Yolanda Ritonga (2024) “test set adalah kumpulan data yang akan digunakan algoritma untuk membangun model trial”.

G. Data Mining

Menurut Budi Aribowo dan Salsa Fairuz (2024:04) “Data mining adalah proses mengeksplorasi dan menemukan pola-pola yang bermanfaat, informasi tersembunyi atau pengetahuan yang dapat diambil dari data yang besar, kompleks, dan seringkali tidak terstruktur”.

Menurut Nugroho, Ma’arif, dan Arrif (2022) “Data mining merupakan suatu proses analisis yang bertujuan untuk menggali informasi berharga, konsisten, dan tersembunyi dari kumpulan data yang besar”.

Untuk memahami lebih dalam tentang proses *Data Mining* dan bagaimana langkah-langkahnya dapat diterapkan secara efektif, penting untuk mengenal komponen-komponen utama yang menjadi fondasi dari teknik ini. Berikut adalah komponen utama dari *Data Mining*:

1. Prapemrosesan Data

Tahap awal dalam proses *Data Mining* adalah melakukan prapemrosesan data. Tahapan ini meliputi pembersihan data dari nilai yang hilang atau tidak valid, penanganan *outlier*, perubahan format data, serta pemilihan fitur yang relevan untuk dianalisis.

2. Eksplorasi Data

Setelah data selesai diproses, langkah berikutnya adalah mengeksplorasi data untuk memahami struktur, pola, serta hubungan yang mungkin ada di dalamnya. Langkah ini melibatkan penggunaan teknik visualisasi data, analisis statistik deskriptif, dan metode eksplorasi lainnya.

3. Pembuatan Model

Langkah selanjutnya setelah eksplorasi adalah membangun model atau algoritma yang sesuai untuk mengungkap pola tersembunyi di dalam data. Tahap ini melibatkan pemilihan model, penentuan parameter, serta pelatihan model menggunakan data yang tersedia.

4. Pengenalan Pola

Setelah model berhasil dibuat dan dilatih, langkah berikutnya adalah menjalankan model tersebut untuk menemukan pola-pola yang bermanfaat dalam data. Pola ini dapat berupa korelasi, klasifikasi, regresi, atau asosiasi yang memberikan wawasan yang berguna.

5. Interpretasi dan Evaluasi

Hasil dari proses Data Mining perlu diinterpretasikan agar dapat dipahami makna dan dampaknya. Evaluasi dilakukan untuk memastikan kualitas dan validitas hasil, serta memastikan model dapat diterapkan pada data baru dengan akurasi yang baik.

6. Penggunaan Pengetahuan

Pengetahuan yang diperoleh dari *Data Mining* dimanfaatkan untuk memperbaiki pengambilan keputusan, meningkatkan efisiensi bisnis, serta mengoptimalkan kinerja sistem secara keseluruhan.

H. Machine Learning

Menurut Lailil Mufikah, Wayan Firdaus M., dan Diva Kurnianingtyas (2023:4)“Definisi lain dari machine learning adalah bidang kecerdasan buatan untuk memperoleh pengetahuan secara otomatis untuk menciptakan suatu sistem yang cerdas.”.

Sementara itu, menurut Yessy Asri, et al. (2024:12)dalam buku berjudul *Machine Learning & Deep Learning: Analisis Sentimen Menggunakan Ulasan Pengguna Aplikasi* menyatakan bahwa “machine learning adalah mesin yang dikembangkan untuk bisa belajar dengan sendirinya tanpa arahan dari penggunanya”.

Dari kedua pengertian diatas dapat diambil kesimpulan bahwa *Machine Learning* adalah metode dalam bidang kecerdasan buatan yang memungkinkan komputer untuk belajar dan memperoleh pengetahuan secara otomatis dari data historis tanpa perlu arahan langsung dari penggunanya. Berdasarkan definisi yang telah disampaikan, *machine learning* bertujuan untuk menciptakan sistem cerdas yang mampu mengekstraksi informasi dari data secara efektif dan mencapai kinerja optimal melalui proses pelatihan model. Pendekatan ini menjadi salah satu pilar penting dalam pengembangan teknologi modern untuk berbagai aplikasi analitik dan pengambilan keputusan.

I. *Supervised Learning*

Menurut Sirfatullah Alfarizi et al (2023) Supervised Learning merupakan sebuah algoritma yang dirancang untuk menggunakan data yang sudah dilabeli untuk mempelajari hubungan antara input dan output.

Beberapa algoritma supervised learning yang populer antara lain:

1. *Regresi linear* merupakan algoritma yang digunakan untuk memprediksi nilai keluaran berdasarkan masukan tertentu dengan memanfaatkan model linier yang diperoleh dari data pelatihan.
2. *Decision tree* adalah algoritma yang membentuk struktur pohon berdasarkan data pelatihan, yang digunakan untuk melakukan proses klasifikasi.
3. *Support Vector Machine (SVM)* merupakan algoritma yang menciptakan model untuk klasifikasi dengan memaksimalkan margin antar kelas yang berbeda.
4. *Naïve Bayes* adalah algoritma klasifikasi berbasis probabilitas yang didasarkan pada konsep teorema Bayes. Algoritma ini menggunakan probabilitas kondisional untuk melakukan prediksi atau klasifikasi data. Meskipun algoritma ini sederhana, *Naïve Bayes* sering kali efektif dan cepat dalam menangani tugas klasifikasi, terutama untuk dataset yang memiliki banyak fitur.

J. *Unsupervised Learning*

Menurut Yessy Asri, et al. (2024:14) “Unsupervised learning merupakan jenis machine learning yang menggunakan data tanpa label untuk mempelajari struktur data”

Menurut Yolanda, Hersyah, dan Marozi (2021) “Unsupervised Learning adalah algoritma yang bertujuan untuk menarik kesimpulan dari dataset”.

Dari kedua pengertian diatas dapat diambil kesimpulan bahwa unsupervised learning adalah algoritma yang belajar dengan dataset tanpa label, untuk mempelajari karakteristik struktur data.

Unsupervised learning belajar dengan cara melakukan clustering yaitu belajar hanya dari karakteristik dataset, dikarenakan dataset yang digunakan tidak mempunyai label seperti supervised learning.

Beberapa algoritma *unsupervised learning* yang populer antara lain:

1. *K-means clustering* adalah algoritma yang digunakan untuk mengelompokkan data ke dalam sejumlah kelompok (k) yang berbeda dengan cara meminimalkan jarak antara setiap data dan titik pusat kelompoknya.
2. *Principal Component Analysis (PCA)* merupakan algoritma yang bertujuan untuk mendekomposisi data menjadi beberapa komponen utama guna menyederhanakan struktur data serta mengurangi jumlah dimensinya.

K. Reinforcement Learning

Menurut Yessy Asri (2024:14) Reinforcement learning adalah algoritma yang belajar dari interaksi dengan lingkungannya, dengan cara mencoba dan memperbaiki tindakan yang diambil untuk mencapai tujuan tertentu.

Menurut Sirfatullah Alfarizi et al (2023)“Reinforcement Learning adalah algoritma yang mengumpulkan pengetahuan (sinyal penguatan) untuk memilih tindakan yang mengarah ke hasil tertinggi yang diharapkan”.

Berdasarkan kedua definisi tersebut, Reinforcement Learning dapat disimpulkan sebagai sebuah algoritma pembelajaran yang berorientasi pada interaksi dengan lingkungan. Algoritma ini bekerja dengan mencoba berbagai tindakan dan memperbaikinya berdasarkan umpan balik berupa sinyal penguatan, sehingga dapat memilih tindakan yang menghasilkan nilai atau hasil yang paling optimal untuk mencapai tujuan tertentu.

L. Deep Learning

Menurut Sirfatullah Alfarizi et al (2023) *Deep Learning* merupakan cabang dari ilmu *Machine Learning* yang algoritmanya terinspirasi oleh struktur otak manusia.

Menurut Yessy Asri (2024:20) “*Deep learning* adalah sub-bidang berdasarkan *Machine Learning* yang bekerja menggunakan atau merepresentasikan data pada dalam lapisan pembelajaran supaya lebih bermakna”.

Menurut Wahyudi Setiawan et al (2020:2)“*Deep Learning* merupakan bagian dari *machine learning* yang melakukan pembelajaran secara hirarki”.

Berdasarkan berbagai definisi yang telah dikemukakan, dapat disimpulkan bahwa *Deep Learning* merupakan cabang dari *Machine Learning* yang bekerja dengan cara merepresentasikan data dalam beberapa lapisan pembelajaran secara hierarkis. Pendekatan ini memungkinkan sistem untuk mengekstraksi fitur secara otomatis dari data yang diberikan, sehingga menghasilkan pemahaman yang lebih mendalam dan bermakna. Selain itu, algoritma *Deep Learning* terinspirasi oleh struktur dan cara kerja otak manusia, yang terdiri dari jaringan saraf tiruan (*Artificial Neural Networks*), sehingga mampu mengolah informasi secara kompleks dan meningkatkan akurasi dalam berbagai tugas kecerdasan buatan..

Deep learning juga di kenal sebagai jaringan saraf tiruan yang memiliki lebih dari 2 lapisan (Hansel, dan Bunyamin, 2021). Struktur tersebut dikenal dengan istilah *Artificial Neural Networks (ANN)* Selain itu, *Deep Learning* memiliki berbagai jenis algoritma, di antaranya adalah :

1. *Convolutional Neural Network (CNN)*
2. *Long Short Term Memory Network (LTSM)*,
3. *Recurrent Neural Network (RNN)*,
4. *Self Organizing Maps (SOM)*.

M. *Convolutional Neural Network (CNN)*

Menurut Pulung Adi Nugroho, Indah Fenriana, dan Rudy Arijanto (2020) “Convolutional Neural Network adalah neural network yang didesain untuk mengolah data dua dimensi”.

Menurut Hidayat, Latifah, dan Damsi (2022) *Convolutional Neural Network (CNN)* adalah salah satu algoritma *deep learning* yang dirancang khusus untuk menganalisis data spasial.

Dari kedua pengertian di atas, dapat disimpulkan bahwa CNN adalah algoritma *deep learning* yang dirancang khusus untuk memproses data spasial, seperti citra, dengan memanfaatkan hubungan korelasi antar neuron untuk mengekstrak fitur-fitur penting secara efisien dan menghasilkan kinerja yang lebih baik dibandingkan metode tradisional.

Algoritma ini bekerja dengan memanfaatkan pembacaan karakteristik data sebagai *input*, yang kemudian direpresentasikan dalam bentuk *tensor*. Pada CNN, setiap neuron dan *perceptron* memiliki hubungan yang saling berkorelasi, memungkinkan algoritma ini untuk mengidentifikasi keterkaitan antar karakteristik secara efisien. Selain itu, CNN mampu memproses data citra secara simultan, sehingga menghasilkan kinerja yang lebih optimal dibandingkan metode konvensional.

Secara umum, CNN terdiri dari beberapa lapisan utama, yaitu:

1. Convolutional Layer

Lapisan ini bertanggung jawab untuk mengekstraksi fitur dari citra input menggunakan filter (kernel). Filter ini akan melakukan operasi konvolusi dengan menggeser dan mengalikan bobotnya terhadap matriks input.

Hasil dari konvolusi ini adalah feature map yang menunjukkan fitur-fitur penting dalam citra, seperti tepi, tekstur, atau pola tertentu.

2. Activation Function (ReLU)

Fungsi aktivasi Rectified Linear Unit (ReLU) digunakan setelah convolutional layer untuk meningkatkan non-linearitas dalam jaringan. Fungsi ini menggantikan nilai negatif dalam feature map dengan nol sehingga model dapat belajar lebih cepat.

3. Pooling Layer

Lapisan ini bertugas untuk mengurangi dimensi feature map dengan mempertahankan informasi yang paling relevan. Terdapat beberapa jenis pooling:

- a. Max Pooling: Memilih nilai maksimum dalam jendela tertentu.
- b. Average Pooling: Menghitung rata-rata dari nilai dalam jendela tertentu.

Pooling membantu mengurangi ukuran data dan meningkatkan efisiensi komputasi serta mengurangi risiko overfitting.

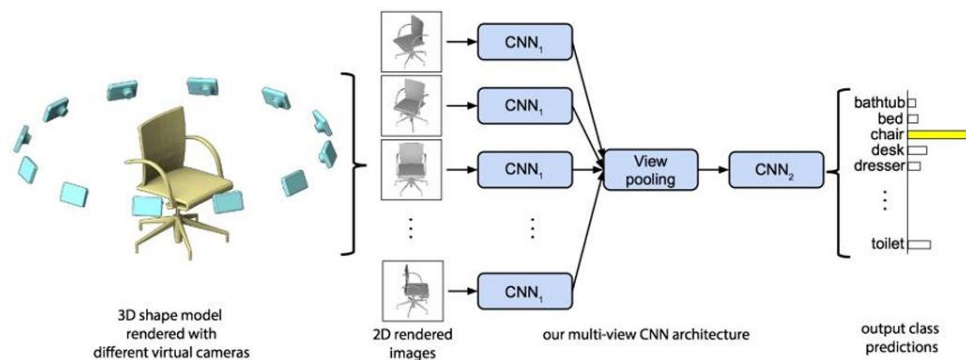
4. Fully Connected Layer (FC Layer)

Setelah serangkaian convolutional dan pooling layer, fitur-fitur yang telah diekstraksi akan dipetakan ke dalam lapisan fully connected. Lapisan ini bertindak seperti jaringan saraf biasa yang menghubungkan semua neuron dan menghasilkan prediksi akhir.

5. Softmax Layer (Output Layer)

Untuk tugas klasifikasi, CNN biasanya menggunakan fungsi aktivasi Softmax yang mengubah output menjadi probabilitas untuk setiap kelas. Probabilitas tertinggi akan menjadi hasil prediksi model.

Jadi *convolutional neural network* adalah sebuah algoritma yang menyerupai neuron otak manusia, yang dapat menganalisis data spasial. CNN memiliki filter-filter yang dilatih untuk tujuan spesifik, misalnya pada lapisan filter pertama ditujukan untuk menganalisis tepian gambar atau perbedaan warna, lapisan berikutnya biasanya lebih diarahkan untuk mempelajari fitur-fitur dari *input object*, dan untuk lapisan terakhir digunakan untuk mengidentifikasi objek.



Gambar 2. 1
Ilustrasi Ekstraksi Fitur CNN pada object 3D ke 2D

Sumber: Cholissodin, I., et al., (2020)

N. ConvNeXtBase

Menurut Liu et al. (2022), ConvNeXtBase adalah model ConvNet modern yang dimodifikasi dari arsitektur ResNet untuk menyaingi Vision Transformers dalam tugas pengenalan gambar, deteksi objek, dan segmentasi.

Dari pengertian diatas dapat disimpulkan bahwa, ConvNeXtBase adalah arsitektur jaringan saraf konvolusional (*Convolutional Neural Network* atau CNN) yang dirancang untuk tugas klasifikasi citra.

Model ini merupakan evolusi dari arsitektur ResNet, dengan beberapa modifikasi untuk meningkatkan kinerja dan efisiensi. ConvNeXtBase dapat digunakan sebagai model dasar (*backbone*) dalam membangun model yang lebih kompleks untuk berbagai kasus penggunaan.

O. EfficientNetV2S

Menurut Tan & Le (2021) EfficientNetV2S adalah salah satu model dalam keluarga EfficientNetV2 yang dirancang untuk meningkatkan efisiensi dan kecepatan pelatihan dalam tugas klasifikasi gambar.

Model ini menggunakan kombinasi pencarian arsitektur jaringan yang mempertimbangkan efisiensi pelatihan dan skala untuk mengoptimalkan kecepatan pelatihan dan efisiensi parameter. EfficientNetV2S dapat digunakan sebagai model dasar (*backbone*) dalam membangun model yang lebih kompleks untuk berbagai kasus penggunaan.

P. *Augmentasi data*

Menurut Hansel, dan Bunyamin (2021) Augmentasi data adalah teknik untuk mengurangi *overfitting* pada pelatihan model *machine learning*. Teknik ini juga dapat meningkatkan akurasi model dengan memodifikasi gambar sedemikian rupa sehingga komputer mengenali gambar tersebut adalah gambar yang berbeda, sementara manusia tetap dapat mengenali bahwa gambar tersebut sama.

Menurut Auliaddina & Arifin (2024) “Augmentasi data adalah cara yang dilakukan untuk secara signifikan menambah keragaman data gambar yang digunakan dengan beberapa Teknik”.

Dari kedua pengertian di atas, dapat disimpulkan bahwa augmentasi data adalah teknik yang digunakan untuk meningkatkan keragaman data, khususnya data gambar, dengan melakukan modifikasi atau transformasi pada gambar tersebut. Tujuannya adalah untuk mengurangi *overfitting*, meningkatkan akurasi model, dan memastikan bahwa model dapat mengenali pola-pola yang lebih umum dari data yang beragam.

Q. *Overfitting*

Menurut Nugroho Adi, Fenriana, dan Arijanto (2020) *Overfitting* adalah kondisi di mana model *machine learning* mendapatkan *loss* yang rendah saat pelatihan (*training*), tetapi mengalami *loss* yang tinggi ketika melakukan klasifikasi pada data baru (prediksi). Penyebab dari *overfitting* dapat bervariasi, di antaranya:

1. Model terlalu kompleks dibandingkan dengan kebutuhan, sehingga terlalu banyak mempelajari detail dari data pelatihan.
2. Dataset terlalu sedikit, sehingga model tidak mendapatkan cukup variasi untuk memahami pola secara umum.
3. Data pelatihan tidak mencakup representasi yang cukup untuk berbagai kondisi atau kasus nyata (*underrepresented data*).
4. Tidak ada atau kurangnya penerapan teknik regulasi seperti *dropout*, *weight decay*, atau *batch normalization*.
5. Proses pelatihan berlangsung terlalu lama (*overtraining*), sehingga model mulai menghafal data pelatihan alih-alih memahami pola.

Untuk menangani *overfitting*, terdapat beberapa metode yang dapat diterapkan, antara lain:

1. Menambah jumlah data pelatihan, Dengan memperluas dataset, model memiliki lebih banyak variasi pola untuk dipelajari, sehingga dapat meningkatkan kemampuan generalisasi.
2. Menggunakan teknik augmentasi data, Teknik ini memodifikasi data pelatihan, seperti rotasi, pemotongan, atau perubahan warna, untuk menciptakan variasi tambahan pada dataset.
3. Menerapkan regulasi, Teknik seperti *L1/L2 regularization* atau *dropout* dapat digunakan untuk mencegah model menjadi terlalu kompleks.
4. Menyederhanakan arsitektur model, Mengurangi jumlah layer atau neuron pada model dapat membantu mengurangi risiko model terlalu mempelajari detail data pelatihan.
5. Menggunakan *early stopping*, Menghentikan proses pelatihan ketika performa pada data validasi mulai menurun, untuk mencegah model terlalu menghafal data pelatihan.
6. Membagi dataset dengan baik. Pastikan data dibagi dengan proporsi yang tepat untuk pelatihan, validasi, dan pengujian, sehingga model dapat dievaluasi dengan benar.
7. Menggunakan teknik *batch normalization*, Metode ini membantu menstabilkan proses pelatihan dan meningkatkan kemampuan generalisasi model.

R. *Underfitting*

Menurut Hansel, dan Bunyamin (2021) *underfitting* adalah kondisi di mana model *machine learning* gagal untuk mempelajari pola yang cukup dari data pelatihan, sehingga performanya buruk pada data pelatihan maupun data baru.

Penyebab dari *underfitting* dapat bervariasi, di antaranya:

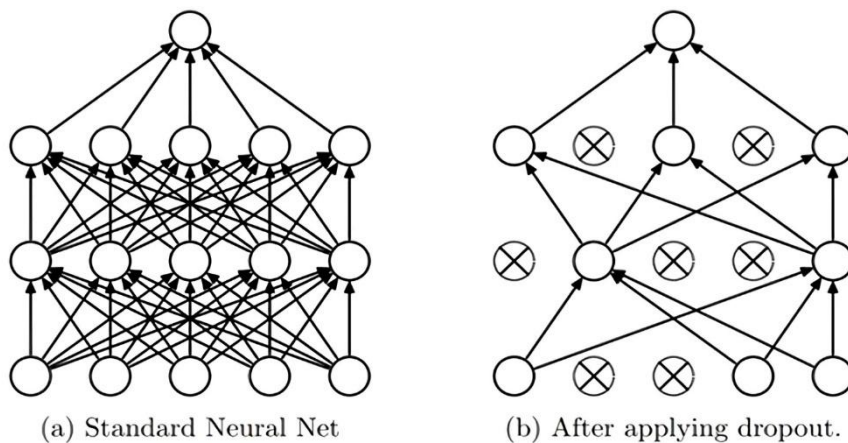
1. Model terlalu sederhana sehingga tidak mampu menangkap pola kompleks dari data.
2. Data pelatihan tidak mencukupi atau tidak memiliki representasi pola yang memadai.
3. Waktu pelatihan terlalu singkat, sehingga model belum mencapai titik konvergensi.
4. Fitur yang digunakan tidak relevan atau kurang memadai untuk merepresentasikan masalah yang ingin diselesaikan.
5. Pemilihan parameter model tidak sesuai, seperti *learning rate* terlalu tinggi yang menyebabkan model melewati pola optimal.

Untuk mengatasi *underfitting*, beberapa langkah yang dapat dilakukan adalah:

1. Menambahkan jumlah layer atau neuron pada arsitektur model untuk menangkap pola yang lebih kompleks.
2. Memastikan model dilatih hingga mencapai titik konvergensi, di mana loss sudah stabil dan tidak mengalami perubahan signifikan.
3. Menggunakan teknik *feature selection* untuk memilih fitur yang benar-benar berkontribusi terhadap prediksi model.
4. Menyediakan lebih banyak data yang representatif agar model memiliki lebih banyak pola untuk dipelajari.
5. Menggunakan *hyperparameter tuning*, mencoba berbagai kombinasi parameter seperti *learning rate*, jumlah layer, atau ukuran batch untuk mendapatkan konfigurasi yang optimal.
6. Menggunakan teknik augmentasi data, memperluas variasi data pelatihan dengan melakukan augmentasi, seperti rotasi, pemotongan, atau perubahan warna.
7. Memperbaiki preprocessing data, memastikan data pelatihan telah diproses dengan baik, seperti normalisasi, penghapusan *outlier*, atau *scaling* yang sesuai.

S. Dropout

Menurut Guntoro, Midyanti, dan Hidayati (2022) *Dropout* adalah sebuah teknik regulasi pada jaringan saraf tiruan di mana beberapa neuron dipilih secara acak untuk dinonaktifkan selama proses pelatihan. *Dropout* berfungsi untuk mencegah *overfitting* terjadi selama proses *training* berlangsung.



Gambar 2. 2
Aplikasi Dropout Pada Neural Network

sumber: Hansel et al (2021)

T. Python

Menurut Rahman et al (2023) “Python adalah bahasa pemrograman yang menggunakan interpreter untuk menjalankan kode programnya”. Python digunakan dalam penelitian kali ini dikarenakan mempunyai library yang mendukung untuk melakukan pemrosesan gambar dan membuat machine learning. Library yang digunakan pada penelitian ini adalah:

1. Numpy

Tools diatas digunakan untuk komputasi ilmiah dan matematika tingkat lanjut. Secara keseluruhan, NumPy berkerja untuk menyediakan lingkungan yang kuat untuk analisis data, simulasi ilmiah dan pemodelan matematika dengan python.

2. Pandas

Pandas berfungsi untuk manipulasi dan analisis data. Library ini menyediakan struktur data fleksibel berupa DataFrame yang memungkinkan pengguna untuk memuat, memodifikasi, serta menganalisis data dengan mudah dalam berbagai format. Pandas sangat bermanfaat dalam berbagai operasi seperti pembersihan, penggabungan, pemfilteran, dan agregasi data.

3. Matplotlib

Matplotlib digunakan untuk membuat visualisasi data. *Library* ini memudahkan pengguna membuat berbagai jenis grafik, seperti grafik garis, batang, lingkaran, histogram, dan lainnya. Matplotlib membantu dalam menunjukkan tren, pola, serta distribusi data sehingga analisis lebih mudah dipahami melalui visualisasi yang informatif.

4. Tensorflow

TensorFlow digunakan untuk membangun serta melatih model *machine learning*, khususnya untuk model *deep learning*. *Library* ini merupakan kerangka kerja (*framework*) yang memungkinkan pembuatan berbagai jenis jaringan saraf tiruan (*neural networks*). Dengan TensorFlow, pengembang dapat menentukan, mengoptimalkan, dan menjalankan grafik komputasi untuk berbagai tugas seperti pengenalan gambar, pemrosesan bahasa alami, dan analisis prediktif.

5. Sckit-learn

Scikit-learn adalah *library* yang menyediakan berbagai algoritma pembelajaran mesin, termasuk untuk klasifikasi, regresi, pengelompokan, pengurangan dimensi, dan pemilihan model. *Library* ini dirancang untuk bekerja selaras dengan *library* lain seperti NumPy dan SciPy, sehingga memudahkan pengguna dalam menerapkan analisis dan pembelajaran mesin.

U. Malaria

Menurut Andra Perdana (2021) malaria adalah penyakit menular yang mematikan, yang disebabkan oleh gigitan nyamuk anopheles betina, Infeksi malaria disebabkan oleh protozoa genus plasmodium. ada beberapa spesies *Plasmodium* yang menyebabkan malaria pada manusia, di antaranya:

1. *Plasmodium falciparum*
2. *Plasmodium vivax*
3. *Plasmodium malariae*
4. *Plasmodium ovale*
5. *Plasmodium knowlesi*

V. Endemi

Menurut Buchari Lapau (2017:34) “endemi adalah suatu keadaan di mana frekuensi penyakit atau masalah kesehatan hampir sama sepanjang tahun”.

W. Plasmodium

Menurut Avichena dan Anggriyani (2023), “*Plasmodium* adalah makhluk hidup bersel satu yang termasuk dalam kelompok protozoa”.

X. Darah

Menurut Pribadi et al (2017) Darah adalah jaringan ikat berbentuk cairan yang terdiri dari beberapa komponen utama, yaitu eritrosit (sel darah merah), leukosit (sel darah putih), trombosit (keping darah), dan plasma darah sebagai cairannya.

Selain itu, darah berfungsi sebagai alat transportasi utama dalam tubuh untuk mengangkut berbagai zat penting.

Y. Citra Darah

Menurut Sholihah (2010) citra darah adalah hasil foto darah yang diamati dengan menggunakan mikroskop inferted.

Z. Penelitian Terdahulu

1. Penelitian Algoritma Tradisional.

Pada penelitian yang dilakukan oleh Aisyah dan Anraeni (2022) memanfaatkan algoritma K-Nearest Neighbor (K-NN) untuk klasifikasi citra darah malaria, dengan akurasi sebesar 89%. Meskipun demikian, metode ini dinilai kurang fleksibel untuk dataset yang lebih kompleks.

Penelitian dengan algoritma tradisional lainnya dilakukan oleh Ain Banyal dan Rachman Dayat (2016) yang menggunakan metode Support Vector Machine (SVM). Hasil penelitian menunjukkan akurasi sebesar 73,33%, namun memiliki kekurangan seperti jumlah data yang terlalu sedikit serta kesalahan deteksi akibat pewarnaan preparat yang kurang optimal.

2. Penelitian Berbasis *Deep Learning*.

Pendekatan berbasis deep learning juga menjadi perhatian utama dalam penelitian deteksi malaria. Penelitian yang dilakukan Shekar et al (2022) menggunakan Convolutional Neural Network (CNN) untuk mengekstraksi fitur dari *plasmodium* dengan tiga pendekatan: basic CNN, frozen layer dari model VGG, dan fine-tune model VGG. Hasilnya, basic CNN mencapai akurasi sebesar

99,85%, frozen layer dari VGG mencapai akurasi 99,56%, sementara fine-tune model VGG mencatat akurasi tertinggi, yaitu 99,89%. Temuan ini menunjukkan bahwa metode fine-tune pada model machine learning memberikan hasil yang lebih baik dibandingkan pendekatan lainnya.

Penelitian lainnya oleh Yohannes et al (2020) menerapkan CNN berbasis saliency, menggunakan beberapa varian seperti Region Contrast Saliency (62,67% akurasi), Frequency-Tuned Saliency (90,32% akurasi), Histogram Contrast Saliency (79,06% akurasi), dan Spectral Residual Saliency (50% akurasi). Menariknya, model tanpa fitur saliency justru mencapai akurasi tertinggi, yaitu 95,06%. Hasil ini juga menunjukkan bahwa penghapusan fitur saliency dapat meningkatkan stabilitas performa model.

3. Penelitian Dengan Menggunakan Arsitektur CNN Ringan.

Model ringan berbasis deep learning juga diperkenalkan dalam beberapa penelitian, dengan tujuan mendukung implementasi di daerah dengan sumber daya terbatas.

Pada penelitian yang dilakukan Lydia et al (2019), arsitektur ResNet-34 digunakan untuk mendeteksi malaria, dengan akurasi sebesar 96,91%. Temuan ini mengindikasikan bahwa arsitektur ResNet-34 cukup andal untuk tugas deteksi malaria, meskipun penelitian ini hanya berfokus pada pengaturan learning rate tanpa mengoptimasi arsitektur untuk mendapatkan akurasi terbaik.

Penelitian yang dilakukan oleh Sasongko (2022) menggunakan arsitektur seperti ResNet-101, AlexNet, dan VGG-19, di mana ResNet-101 mencapai akurasi 97,58% dengan penyesuaian learning rate. Selain itu penelitian yang dilakukan Zein pada tahun 2019 juga menggunakan arsitektur ResNet, yang mampu mencapai akurasi 97% pada data pengujian. Pendekatan ini memungkinkan deteksi parasit malaria dengan waktu pelatihan yang singkat tanpa mengorbankan tingkat akurasi .

4. Penelitian Dengan Realtime Object Detection

Penelitian pada realtime object detection juga pernah dilakukan dengan menggunakan arsitektur CNN YOLOv4 dan YOLOv5, pada penelitian ini didapatkan akurasi dari YOLOv4 sebesar 90,51% (Sukumarran et al 2024), dan YOLOv5 sebesar 60,7% (Koirala et al 2022).