

## **BAB II**

### **KERANGKA TEORITIS**

#### **A. Landasan Teori**

##### **1. Artificial Intelligence secara umum**

Artificial Intelligence atau umumnya kecerdasan buatan merupakan kecerdasan atau kepintaran yang dibuat oleh manusia yang diberikan kepada mesin untuk bisa berpikir layaknya seperti manusia. Menurut Siti Nurmaini et al. (2021, pp. 15–17) Kecerdasan buatan (Artificial Intelligence, AI) bisa dikatakan sebagai sebuah disiplin ilmu untuk bisa membuat mesin yang bisa melakukan banyak hal yang memerlukan kecerdasan layaknya yang biasa dilakukan oleh manusia, Ada beberapa penelitian yang melakukan penelitian untuk membuktikan kapan suatu program atau mesin bisa dikatakan cerdas, usaha ini dilakukan matematikawan dari Inggris bernama Alan Turing, beliau merupakan inventor pengujian turing “Turing Test”, yang dapat memutuskan mesin/program tersebut dikatakan cerdas atau tidak, Proses ini melibatkan seorang manusia A, yang menggunakan keyboard dan layar komputer, yang menanyakan pertanyaan pertanyaan yang sama pada komputer dan manusia B. Jika manusia A tidak bisa membedakan jawaban yang diberikan oleh komputer dari jawaban yang diberikan manusia B, Maka bisa dikatakan bahwa komputer (Program/Mesin) itu cerdas. Kecerdasan buatan atau AI ini ternyata telah berkembang pesat dan semakin mendekati kemampuan manusia dalam hal berpikir ditandai dengan pencapaiannya seperti yang dilakukan *Turing Test*;

Teknologi AI ini dimulai dari adanya orang – orang yang berpikiran bagaimana mesin bisa berpikir layaknya manusia. Menurut Neny Rosmawarni et al. (2024, pp. 17–18) AI lahir sekitar tahun 1950an ketika beberapa perintis mulai memiliki pemikiran bagaimana membuat komputer bisa “berpikir”, banyak ahli yang percaya bahwa AI tingkat manusia dapat dicapai dengan membuat seperangkat aturan eksplisit yang cukup agar bisa memanipulasi pengetahuan oleh programmer. AI terus berkembang di sekarang dan bisa menyamai manusia dalam menyelesaikan tugasnya.

##### **2. Machine Learning secara umum**

Machine learning secara umum bisa dikatakan sebagai sebuah mesin yang bisa diberikan materi layaknya manusia untuk bisa dipelajari yang bertujuan untuk bisa membantu pekerjaan manusia. Menurut Neny Rosmawarni (2024, pp. 12–15) mengatakan Machine Learning ini bisa dibidang teknologi kecerdasan buatan yang dirancang untuk meniru atau menggantikan peran manusia dalam melakukan aktivitas dan memecahkan masalah dan juga *Machine Learning* ini adalah mesin

yang didesain untuk bisa belajar dan beroperasi tanpa instruksi apapun dari pengguna, dalam machine learning terbagi ke dalam 3 kategori yaitu *Supervised Learning* (Algoritma dimana datanya berlabel), *Unsupervised Learning* (Algoritma yang datanya tidak berlabel), *Reinforcement Learning* (Algoritma yang tidak dapat digolongkan menjadi supervised learning dan unsupervised learning). Maka berdasarkan pemaparan diatas potensi yang dimiliki oleh machine learning sangatlah besar untuk bisa terus dikembangkan sehingga akan mampu untuk membantu tugas yang tidak bisa dikerjakan oleh manusia;

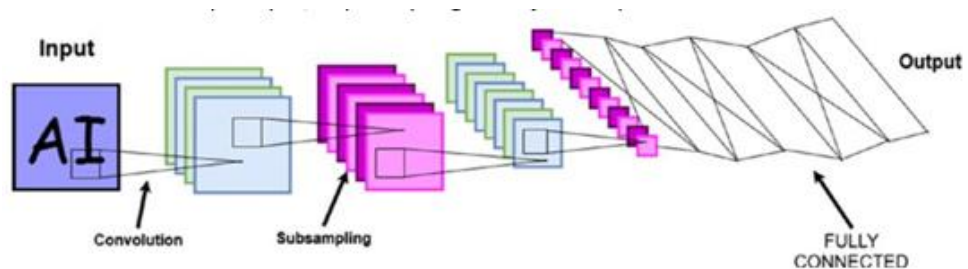
*Deep Learning* (DL) merupakan salah satu bidang dalam kecerdasan buatan atau AI yang semakin berkembang pesat dan banyak digunakan dalam teknologi modern. Menurut Neney Rosmawarni (2024, p. 128) *Deep Learning* merupakan salah satu bagian dari Machine Learning yang terdiri dari algoritma pemodelan abstraksi tingkatan teratas pada data memanfaatkan sekumpulan peranan transformasi non-linear yang ditata berlapis – lapis serta mendalam, DL ini sangat bagus untuk digunakan pada supervised learning, unsupervised learning serta semi-supervised learning ataupun untuk reinforcement learning dalam aplikasi semacam indentifikasi gambar, suara, pengkategorian teks, serta sejenisnya. Kemampuan luar biasa yang dimiliki oleh DL dalam menangani data kompleks dan beragam membuat DL ini memiliki potensi untuk terus berinovasi di masa depan.

### **3. Konsep CNN**

Convolutional Neural Network (CNN) merupakan salah satu algoritma deep learning yang dirancang khusus untuk analisis data visual namun bisa juga digunakan. Budi Raharjo (2022, p. 55) menjelaskan, Convolutional Neural Networks atau Jaringan Saraf Convolutional (CNN) adalah jaringan saraf tiruan feed-forward yang dalam di mana jaringan saraf mempertahankan struktur hierarkis dengan mempelajari representasi fitur internal dan menggeneralisasi fitur dalam masalah gambar umum seperti pengenalan objek dan masalah penglihatan komputer lainnya. Wahyudi Setiawan (2020, p. 8) menjelaskan bahwa CNN mampu mengekstraksi fitur dari gambar dan video dengan menggunakan lapisan konvolusi untuk mendeteksi pola-pola tertentu. Pernyataan ini menunjukkan bahwa CNN memiliki kemampuan kuat dalam memahami struktur visual yang kompleks.

Dalam beberapa kasus Convolutional Neural Network (CNN) lebih mudah digunakan dalam melakukan pendeteksian gambar dibanding metode lain salah satunya Multi-Layer Perceptron MLP. Menurut Aston Zhang et al. (2022, p. 233-234) CNN ini unggul dibandingkan MLP dikarenakan dalam metode MLP dalam melakukan pendeteksian tidak memanfaatkan struktur spasial gambar, dibandingkan dengan CNN yang mampu mengenali pola lokal antar piksel secara efisien, lebih

akurat dan komputasi yang lebih hemat. CNN bisa dibilang merupakan salah satu metode pendeteksian yang cukup mampu mendeteksi objek jauh lebih akurat dan efisien.



**Gambar 2. 1 Arsitektur CNN (Budi Raharjo, 2022)**

Dalam proses pendeteksian objek CNN memanfaatkan struktur alami pada gambar sehingga akan lebih efisien dalam melakukan pendeteksian gambar. Menurut Aston Zhang et al. (2022, pp. 234–238) CNN memiliki kemampuan yang disebut spasial invariance, tetap bisa mengenali objek meskipun letaknya berubah dalam gambar ini menjadi keunggulan yang memungkinkan model memindai seluruh gambar tanpa harus mengetahui posisi pasti objek sejak awal, seperti pada gambar 2.2:



**Gambar 2. 2 Gambar Contoh Kasus CNN (Aston Zhang, 2022)**

Seperti pada gambar diatas ada kasus dalam permainan Where's Waldo? Dalam permainan tersebut diminta untuk bisa mengenali keberadaan waldo, sekilas dalam foto tersebut terlihat sama semua, namun dengan menggunakan CNN maka akan terdeteksi seperti pada gambar 2.3:



**Gambar 2. 3 CNN Mendeteksi Waldo (Aston Zhang, 2022)**

Meskipun objek yang dideteksi berada dalam keramaian CNN tetap mampu mendeteksi objek dengan sangat baik.

Dalam proses membangun model CNN terdapat formula didalamnya seperti contoh untuk gambar 2.3 dalam menemukan objek yang dicari. Menurut Aston Zhang et al. (2022, pp. 235–239) untuk bisa membuat model dalam menemukan object harus melalui tahap tranformasi dari fully connected ke representasi spasial dengan formula seperti pada gambar 2.4:

$$\begin{aligned}
 [H]_{i,j} &= [U]_{i,j} + \sum_k \sum_l [W]_{i,j,k,l} [X]_{k,l} \\
 &= [U]_{i,j} + \sum_a \sum_b [V]_{i,j,a,b} [X]_{i+a,j+b}.
 \end{aligned}$$

**Gambar 2. 4 Formula Fully Connected (Aston Zhang, 2022)**

Kemudian setelah itu akan masuk translation invariance atau disebut dengan konvolusi awal dengan formula seperti pada gambar 2.5:

$$[H]_{i,j} = u + \sum_{a=-\Delta}^{\Delta} \sum_{b=-\Delta}^{\Delta} [V]_{a,b} [X]_{i+a,j+b}.$$

**Gambar 2. 5 Formula Translation Invariance (Aston Zhang, 2022)**

Setelah itu definisi matematis konvolusi secara umum, dengan formula seperti pada gambar 2.6:

$$(f * g)(i, j) = \sum_a \sum_b f(a, b)g(i - a, j - b).$$

**Gambar 2. 6 Formula Konvolusi (Aston Zhang, 2022)**

Terakhir ada tahap konvolusi multi-channel, dengan formula seperti pada gambar 2.7:

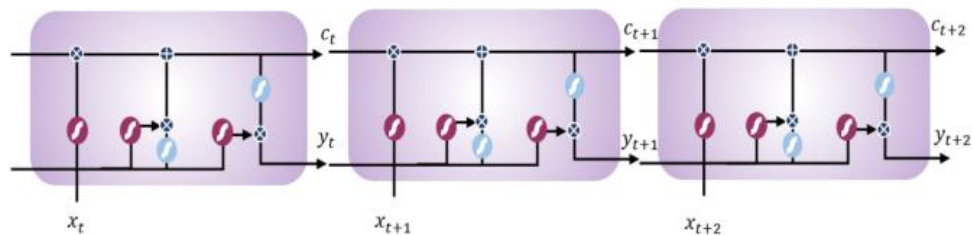
$$[H]_{i,j,d} = \sum_{a=-\Delta}^{\Delta} \sum_{b=-\Delta}^{\Delta} \sum_c [V]_{a,b,c,d} [X]_{i+a,j+b,c},$$

**Gambar 2. 7 Formula Konvolusi Multi-Channel (Aston Zhang, 2022)**

CNN dibuktikan mampu dalam hal mengekstraksi representasi fitur kompleks dari input berwarna maupun yang bertingka.

#### 4. Konsep LSTM

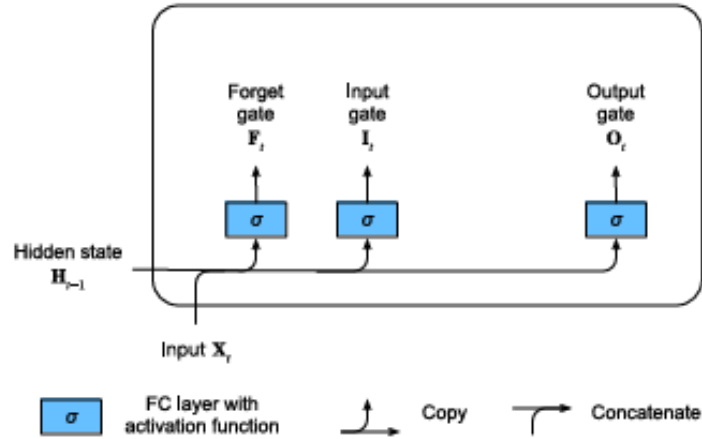
Menurut Yudistira et al. (2023, pp. 48–49), istilah “Long Short-Term Memory” mencerminkan kemampuan jaringan untuk menyimpan informasi dalam jangka panjang melalui bobot, serta dalam jangka pendek melalui aktivasi sementara yang berpindah dari satu node ke node berikutnya. Yudistira et al. (2023, pp. 49) menambahkan, model LSTM memperkenalkan penyimpanan tambahan melalui sel memori, yang merupakan unit komposit berisi node dengan pola konektivitas tertentu dan melibatkan operasi matematis seperti perkalian. Struktur ini memungkinkan LSTM menyimpan dan mengelola informasi secara efisien dalam konteks data berurutan yang kompleks.



**Gambar 2. 8 Arsitektur LSTM (Yudistira et al, 2023)**

Yudistira et al (2023, pp. 49) menjelaskan bahwa model LSTM menyerupai jaringan saraf rekuren standar dengan lapisan tersembunyi (*hidden layer*), tetapi setiap node pada lapisan tersembunyi digantikan oleh sel memori. Menurut Yudistira et al. (2023, pp. 49), setiap sel memori berisi simpul dengan tepi berulang yang terhubung sendiri dengan bobot tetap, memastikan bahwa gradien tidak menghilang atau terlalu besar nilainya setelah melewati beberapa langkah. Aston Zhang et al. (2022, p. 371) menjelaskan, LSTM menggunakan sel memori yang dilengkapi tiga gerbang utama yaitu *input gate* yang mengatur seberapa banyak informasi baru

yang akan disimpan ke dalam *cell state*, *forget gate* yang mengatur seberapa banyak informasi lama yang akan dibuang dari *cell state*, dan *output gate* yang mengatur seberapa banyak informasi dari *cell state* yang akan dikeluarkan sebagai *hidden state*.



**Gambar 2. 9 Sel Memori LSTM (Aston Zhang, 2022)**

Secara matematis, misalkan terdapat  $h$  unit tersembunyi (*hidden units*), ukuran *batch* adalah  $n$ , dan jumlah input adalah  $d$ . Aston Zhang et al. (2022, p. 371) menjelaskan, input adalah  $X_t \in \mathbb{R}^{n \times d}$  dan *hidden state* dari langkah waktu sebelumnya adalah  $H_{t-1} \in \mathbb{R}^{n \times h}$ . Sehubungan dengan itu, gerbang-gerbang pada langkah waktu  $t$  didefinisikan sebagai berikut, gerbang input adalah  $I_t \in \mathbb{R}^{n \times h}$ , gerbang lupa adalah  $F_t \in \mathbb{R}^{n \times h}$ , dan gerbang output adalah  $O_t \in \mathbb{R}^{n \times h}$  (Aston Zhang et al., 2022, p. 371). Mereka dihitung sebagai gambar 2.2 berikut:

$$\begin{aligned} I_t &= \sigma(X_t W_{xi} + H_{t-1} W_{hi} + b_i), \\ F_t &= \sigma(X_t W_{xf} + H_{t-1} W_{hf} + b_f), \\ O_t &= \sigma(X_t W_{xo} + H_{t-1} W_{ho} + b_o), \end{aligned}$$

**Gambar 2. 10 Formula Gerbang LSTM (Aston Zhang, 2022)**

Di mana  $W_{xi}, W_{xf}, W_{xo} \in \mathbb{R}^{d \times d}$  dan  $W_{xhi}, W_{hf}, W_{ho} \in \mathbb{R}^{h \times h}$  adalah parameter bobot, serta  $b_i, b_f, b_o \in \mathbb{R}^{1 \times h}$  adalah parameter bias (Aston Zhang et al., 2022, p. 371).

Sebelum menentukan bagaimana gerbang mempengaruhi memori, kita definisikan terlebih dahulu *input node*  $\tilde{C}_t \in \mathbb{R}^{n \times h}$  yang dihitung menggunakan fungsi aktivasi tanh:

$$\tilde{C}_t = \tanh(X_t W_{xc} + H_{t-1} W_{hc} + b_c),$$

Selanjutnya, *cell state* internal  $C_t \in \mathbb{R}^{n \times h}$  diperbarui menggunakan operasi Hadamard (perkalian elemen per elemen) sebagai berikut:

$$C_t = F_t \odot C_{t-1} + I_t \odot \tilde{C}_t$$

Di sini, forget gate menentukan seberapa banyak informasi lama yang dipertahankan, sedangkan input gate mengatur seberapa banyak informasi baru yang ditambahkan ke memori. Kombinasi ini memungkinkan LSTM untuk mempertahankan informasi penting dalam jangka panjang dan memperbarui memori secara selektif.

Terakhir, *hidden state*  $H_t \in \mathbb{R}^{n \times h}$  yang menjadi output dari LSTM dihitung dengan menggabungkan cell state dan output gate:

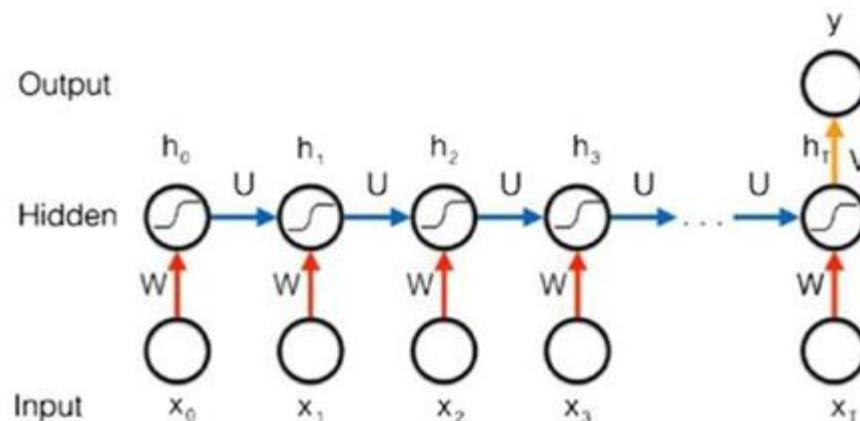
$$H_t = O_t \odot \tanh(C_t)$$

Output gate berfungsi sebagai filter yang mengatur seberapa besar pengaruh memori internal terhadap lapisan berikutnya. Ketika nilai output gate mendekati 1, informasi dari memori internal disalurkan secara penuh; sebaliknya, jika mendekati 0, pengaruhnya diminimalkan.

Budi Raharjo (2022, p. 70) menjelaskan bahwa LSTM dapat memiliki salah satu dari mode berikut:

- (a) Model satu lawan satu;
- (b) Model satu ke banyak;
- (c) Model banyak ke satu;
- (d) Model banyak ke banyak.

Berdasarkan penjelasan tersebut, penelitian ini menerapkan model banyak-ke-satu (*many-to-one*), sebagaimana ditunjukkan pada Gambar 2.11. Pada model ini, sekumpulan frame video yang telah diekstraksi fiturnya oleh CNN disusun sebagai urutan input untuk LSTM, kemudian diproses untuk menghasilkan satu keluaran berupa prediksi kelas akhir.

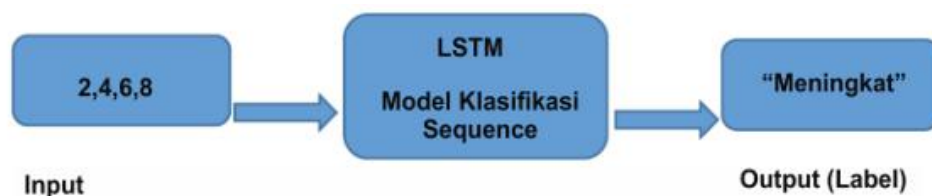


**Gambar 2. 11 Model LSTM Many to one (Budi Raharjo, 2022)**

Budi Raharjo (2022, p. 71) menjelaskan bahwa menjelaskan bahwa *Long Short-Term Memory* (LSTM) sangat cocok digunakan untuk memproses data



berbentuk urutan (*sequence data*) karena kemampuannya dalam melakukan prediksi, klasifikasi, maupun menghasilkan data urutan baru. Contoh urutan adalah rangkaian pengujian di mana stempel waktu dan nilai berada dalam urutan (kronologis) dari urutan, contoh lain adalah video, yang dapat dianggap sebagai rangkaian gambar atau rangkaian klip audio. Budi Raharjo (2022, p. 71) mengatakan bahwa salah satu bentuk penerapan LSTM adalah klasifikasi urutan (*sequence classification*), yaitu proses memberikan label atau kategori pada keseluruhan urutan data berdasarkan pola yang terkandung di dalamnya. Dalam konteks penelitian ini, klasifikasi urutan dilakukan dengan memanfaatkan LSTM untuk menganalisis rangkaian fitur spasial yang diekstraksi dari setiap frame video oleh CNN, sehingga model dapat menentukan kelas akhir, berdasarkan pola temporal yang terbentuk di sepanjang urutan frame tersebut.

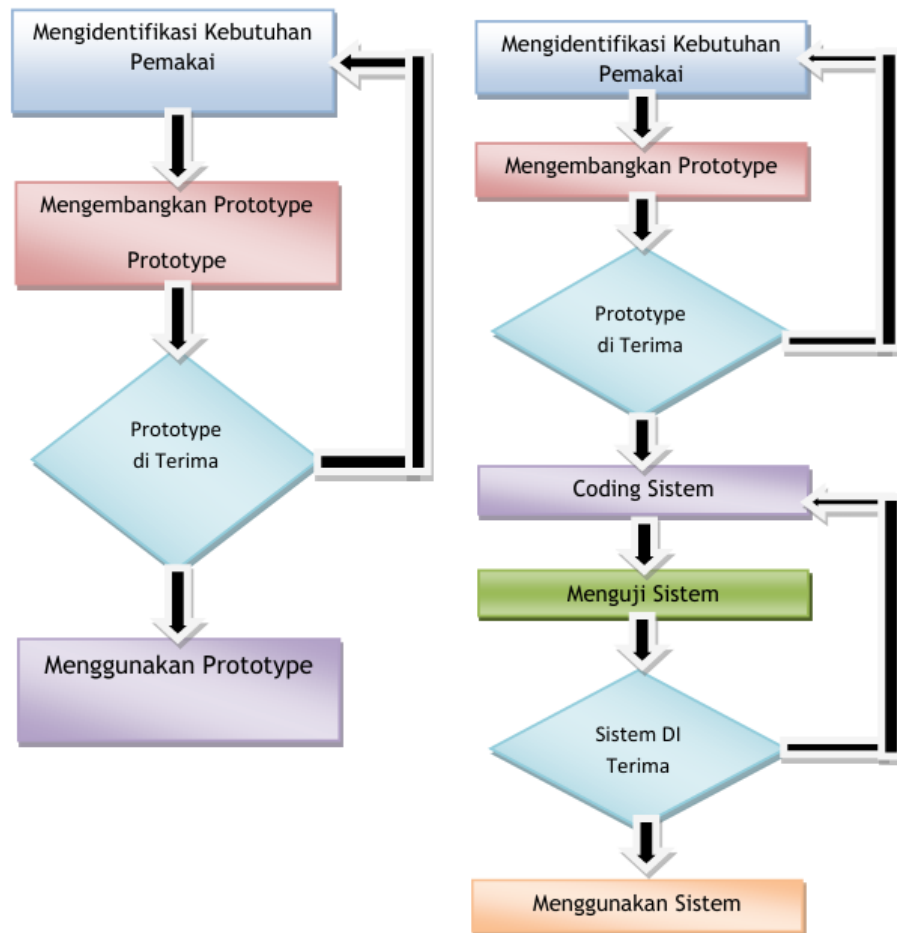


**Gambar 2. 12 Contoh Model Klasifikasi LSTM (Budi Raharjo, 2022)**

## 5. Penerapan Prototyping

Prototyping ini merupakan rancangan awal dari pengembangan suatu produk atau software yang menjadi tolak ukur dalam pembuatan suatu produk atau software. Menurut Budi Hartono (2021, p. 71) prototyping merupakan pengembangan dari sistem dimana permintaan yang ada diubah atau dibuat ke dalam sistem yang bekerja yang akan bekerja secara terus menerus akan diperbaiki dengan kerjasama antara user dan analisis, Prototyping bisa dibangun dengan menggunakan beberapa tool untuk bisa menyederhanakan proses, Alur dari proses prototyping seperti gambar 2.2:





**Gambar 2. 13 Jenis Konsep Prototyping (Budi Hartono, 2021, p. 72)**

Digunakannya proses prototyping ini tidak akan terlepas dari kelebihan dan juga kekurangan didalamnya. Menurut Budi Hartono (2021, p. 72) dalam proses prototyping kelebihan yang dimilikinya adalah dilibatkannya user dalam proses analisa dan juga pengembangan sehingga sistem yang dibuat akan sangat sesuai dengan kebutuhan namun di sisi lain kekurangannya adalah kurang fleksibel dalam menghadapi perubahan dan tidak mudah untuk langsung dirubah desain prototypingnya. Penggunaan prototype perlu diperhitungkan secara matang karena ini akan sangat menentukan pengembangan sistem kedepannya

## 6. Proses Evaluasi Model

Pada saat model AI dikembangkan maka akan melalui proses penilaian kinerja dengan cara kuantitatif. Menurut Wira and Putra (2020, p. 46) umumnya pengujian kinerja akan dinilai menggunakan metode kuantitatif, namun juga bisa melakukan evaluasi model secara eksternal dengan cara melewati data pada model jadi akan langsung melakukan evaluasi terhadap final modelnya dengan

menggunakan performance measure. Pengujian kinerja model ini umumnya menggunakan cara kuantitatif dan bisa dilakukan secara eksternal.

Setiap model AI yang telah perlu melalui tahapan evaluasi untuk bisa melihat apakah sudah sesuai dengan ketentuan dan kebutuhan. Menurut Wira and Putra (2020, p. 47) menjelaskan dalam proses evaluasi model ada namanya Data Splitting, didalamnya terdapat proses training, validation dan juga testing data dalam training data ini digunakan untuk bisa melatih model, validation untuk mengukur kinerja model dan testing mengevaluasi kinerja model. Ada tentang Overfitting dan Underfitting, berkaitan dengan karakteristik dari model bisa model dikatakan terlalu cocok ataupun buruk dan terakhir ada Cross Validation berkaitan dengan teknik untuk menguji kemampuan generalisasi model dengan cara membagi data menjadi beberapa subsampel. Dalam tahap evaluasi model ada beberapa poin yang harus dilakukan Data Splitting, Overfitting dan Underfitting dan Cross Validation, data tidak akan berhenti pada saat training tapi pada saat data dan performa model juga akan dievaluasi.

## **7. Konsep Research and Development**

Penelitian dan pengembangan bisa berupa teknologi atau bukan dalam bentuk teknologi. Menurut Sugiyono (2024, p. 394) mengemukakan bahwa dalam penelitian dan pengembangan merupakan proses atau metode untuk memvalidasi dan mengembangkan produk, bisa berupa buku teks, film ataupun pembelajaran dan juga software. Dari pemaparan diatas dapat dipahami bahwa penelitian dan pengembangan berorientasi untuk menghasilkan produk tentu.

Pengujian dengan metode *One-Group Pretest-Posttest Design* merupakan salah satu cara yang digunakan untuk mengetahui efektivitas dari proses yang terjadi sebelum dan sesudah menggunakan sistem. Menurut Sugiyono (2024, pp. 112–114) metode *one-group pretest-posttest design* merupakan metode yang dimana dinilai nilai observasi yang dilambangkan dengan O1 (nilai pretest), O2 (nilai posttest) dan juga X (Treatment yang diberikan). Dengan menggunakan metode pretest test itu dapat mengetahui perbandingan efektivitas sebelum dan sudah menggunakan sistem.

## **8. Konsep Blackbox dan Whitebox Testing**

Pengujian whitebox merupakan pengujian yang dilakukan untuk menguji dan melihat struktur internal code. Menurut Soetam and Wicaksono (2021, pp. 20–21) pengujian whitebox berfokus pada struktur internal atau alur kerja sebuah sistem atau aplikasi yang sedang dikembangkan. Digunakannya pengujian whitebox bisa melakukan uji terhadap pemenuhan dari spesifikasi sistem.

Pengujian blackbox ini merupakan salah satu cara yang biasanya digunakan untuk bisa mengukur fungsionalitas dari suatu sistem. Menurut Soetam and Wicaksono (2021, pp. 37–41) pengujian blackbox testing sering disebut pengujian fungsional merupakan sebuah metode yang digunakan untuk menguji software yang akan digunakan tanpa melihat dari struktur kodenya maupun programnya. Dengan menggunakan blackbox maka pengujian akan berfokus pada output dari inputan citra yang diinputkan.

## **B. Tinjauan Studi**

Penelitian relevan yang memanfaatkan CNN dalam hal deteksi yang sudah pernah dilakukan sebelumnya mulai dari metode – metode computer vision hingga sampai implementasi metode machine learning dan deep learning. Beberapa hasil penelitian berikut akan dijadikan sebagai referensi untuk penelitian yang akan dilakukan. Penelitian yang dimaksud sebagai berikut:

(1) **“A Shallow System Prototype for Violent Action Detection in Italian Public Schools”** (Perseghin and Foresti, 2023)

Penelitian ini mengembangkan prototipe sistem deteksi kekerasan bernama School Violence Detection (SVD) menggunakan arsitektur 2D Convolutional Neural Network (CNN) dan transfer learning. Sistem ini dirancang untuk mendeteksi tindakan kekerasan di lingkungan sekolah dengan perangkat keras berbiaya rendah. Dataset baru, Daily School Break (DSB), berisi video asli siswa Italia yang mencakup adegan kekerasan dan non-kekerasan. Dengan transfer learning menggunakan model VGG19, sistem mencapai akurasi 95%, membuktikan efektivitasnya dalam mendeteksi kekerasan secara otomatis di lingkungan pendidikan.

(2) **“Violence detection in crowd videos using nuanced facial expression analysis”** (G. and Saleem, 2024)

Penelitian ini mengembangkan metode pendeteksian kekerasan dalam video kerumunan menggunakan Convolutional Restricted Boltzmann Machine (CRBM) dan model VGG16. Wajah dideteksi melalui CRBM dengan filter Gabor untuk ekstraksi fitur, kemudian ekspresi wajah seperti marah, cemas, dan takut dikenali menggunakan VGG16. Hasil analisis digunakan dalam model regresi logistik untuk memprediksi probabilitas kekerasan. Sistem ini dievaluasi dengan matriks kebingungan, ROC curve, dan AUC, menghasilkan tingkat akurasi sebesar 88.25%. Pendekatan ini membuktikan efektivitasnya dalam mendeteksi kekerasan dengan menganalisis ekspresi wajah secara mendalam pada video kerumunan.

(3) **“Detecting Violence in Videos using Convolutional Neural Networks”** (Haiura and Iftene, 2024)

Penelitian ini mengembangkan model deteksi kekerasan berbasis 3D Convolutional Neural Network (CNN) untuk mendeteksi tindakan kekerasan dalam video secara real-time. Model ini menggunakan YOLO v8n untuk mendeteksi individu dalam video, mengolah setiap 16 frame menjadi satu stack, lalu mengidentifikasi kekerasan menggunakan arsitektur 3D CNN. Dataset yang digunakan meliputi "Violence in Movies", "Crowd Violence", dan "Hockey Fights", dengan pembagian data untuk pelatihan (70%), validasi (20%), dan pengujian (10%). Model mencapai tingkat akurasi 87.25% saat diuji, menunjukkan efektivitasnya dalam mengidentifikasi pola kekerasan yang kompleks dari berbagai situasi. Pendekatan ini mendukung pengawasan keamanan publik dan membuka peluang pengembangan lebih lanjut dengan integrasi data multisaluran.

(4) **"Violence detection using spatiotemporal features with 3D Convolutional Neural Network"** (Ullah *et al.*, 2019)

Penelitian ini mengembangkan sistem deteksi kekerasan menggunakan arsitektur 3D Convolutional Neural Network (CNN) untuk menganalisis fitur spasial dan temporal dari video pengawasan. Sistem ini terdiri dari tiga tahap: deteksi orang menggunakan MobileNet-SSD, ekstraksi fitur menggunakan 3D CNN, dan klasifikasi kekerasan dengan Softmax. Sistem diuji pada tiga dataset, yaitu Hockey Fight, Violent Crowd, dan Violence in Movies, dengan akurasi masing-masing 96%, 98%, dan 99.9%. Optimisasi menggunakan toolkit OPENVINO memastikan efisiensi pemrosesan, menjadikan sistem ini unggul dalam deteksi kekerasan otomatis untuk berbagai skenario.

(5) **"A CNN-RNN Combined Structure for Real-World Violence Detection in Surveillance Cameras"** (Vosta and Yow, 2022)

Penelitian ini mengembangkan sistem deteksi kekerasan menggunakan kombinasi arsitektur ResNet50 dan ConvLSTM untuk mendeteksi perilaku abnormal pada data video pengawasan. Sistem ini memanfaatkan ResNet50 untuk ekstraksi fitur spasial dan ConvLSTM untuk analisis temporal, diterapkan pada dataset UCF-Crime yang berisi 13 kategori perilaku abnormal, seperti perkelahian, pencurian, dan vandalisme. Dengan pendekatan ini, sistem mencapai AUC sebesar 81,71% untuk klasifikasi biner (normal vs abnormal), mengungguli metode seperti C3D. Studi ini menyoroti pentingnya kombinasi fitur spasial-temporal dalam mendeteksi kekerasan secara otomatis di lingkungan nyata.

(6) **"Low-Cost CNN for Automatic Violence Recognition on Embedded System"** (Vieira *et al.*, 2022)

Penelitian ini mengembangkan sistem deteksi kekerasan berbasis arsitektur Mobile Convolutional Neural Networks (CNNs) yang diadaptasi untuk pengenalan

otomatis tindakan kekerasan dalam video. Dataset terdiri dari adegan kekerasan dan non-kekerasan di lingkungan padat dan tidak padat. Dengan prototipe berbasis Raspberry Pi, sistem mencapai akurasi hingga 92,05% dan kecepatan pemrosesan 4 frame per detik. Sistem ini juga dilengkapi mekanisme peringatan yang mendeteksi perilaku pra-kekerasan, memungkinkan pencegahan lebih dini. Hasilnya menunjukkan efektivitas model CNN mobile untuk implementasi real-time dengan biaya rendah.

- (7) **“Efficient Violence Detection in Surveillance”** (Vijeikis, Raudonis and Dervinis, 2022)

Penelitian ini mengembangkan sistem deteksi kekerasan berbasis video pengawasan menggunakan arsitektur U-Net yang memanfaatkan MobileNet V2 sebagai encoder untuk ekstraksi fitur spasial, diikuti oleh Long Short-Term Memory (LSTM) untuk ekstraksi fitur temporal dan klasifikasi. Model ini dirancang untuk perangkat edge dengan konsumsi daya rendah (<20 W). Dataset yang digunakan meliputi RWF-2000, Movie Fights, dan Hockey Fights. Hasil eksperimen menunjukkan akurasi rata-rata 82% dengan presisi 81% pada dataset kompleks RWF-2000. Model ini mampu mendeteksi kekerasan dalam waktu nyata dengan latensi rendah, menjadikannya efisien untuk aplikasi pengawasan di perangkat edge. Hasilnya menunjukkan bahwa pendekatan ini dapat meningkatkan waktu respons dan mengurangi beban jaringan.

- (8) **“PENERAPAN METODE CONVOLUTIONAL NEURAL NETWORK (CNN) DAN LONG SHORT TERM MEMORY (LSTM) UNTUK PENGENALAN AKTIVITAS MANUSIA PADA CCTV DI AREA TAMBAK UDANG”** (Aghassi Zulfikar, Somantri and Sudjadi, 2021)

Penelitian ini mengembangkan sistem keamanan berbasis CCTV untuk tambak udang dengan memanfaatkan metode Convolutional Neural Network (CNN) dan Long Short-Term Memory (LSTM). Sistem ini mendeteksi dan mengenali aktivitas manusia, seperti berjalan dan berlari, melalui rekaman CCTV dengan tahapan deteksi objek menggunakan YOLO, pelacakan dengan DeepSort, dan pengenalan aktivitas menggunakan kombinasi CNN dan LSTM. Hasil pengujian menunjukkan akurasi rata-rata 79,25% untuk deteksi manusia dan 55% untuk pengenalan aktivitas, serta keberhasilan hampir 100% dalam pengiriman notifikasi ke smartphone dan penyimpanan gambar ke database. Sistem ini dirancang untuk meningkatkan keamanan tambak udang, terutama dalam mencegah kejahatan seperti pencurian pada malam hari.

- (9) **“Kombinasi Deteksi Objek, Pengenalan Wajah dan Perilaku Anomali menggunakan State Machine untuk Kamera Pengawas”** (NURYASIN, MACHBUB and YULIANTI, 2023)

Penelitian ini merancang sistem pengawasan kamera berbasis tiga modul utama, yaitu deteksi objek menggunakan HOG-SVM, pengenalan wajah dengan CNN berbasis arsitektur VGG-16 menggunakan transfer learning, dan deteksi perilaku anomali dengan spatiotemporal autoencoder berbasis threshold. Ketiga modul ini diintegrasikan menggunakan state machine untuk menciptakan sistem pengawasan yang otomatis dan efisien. Hasil pengujian menunjukkan akurasi 88% untuk deteksi objek, 98% untuk pengenalan wajah, dan 78% untuk deteksi perilaku anomali. Sistem dapat diakses secara nirkabel melalui antarmuka berbasis web, menawarkan fleksibilitas tinggi untuk implementasi di berbagai lingkungan pengawasan.

- (10) **“Real-Time Video Violence Detection Using CNN”** (Shaheer and U, 2023)

Penelitian ini mengembangkan sistem deteksi kekerasan berbasis video secara real-time menggunakan arsitektur MobileNet V2 sebagai ekstraktor fitur spasial dan LSTM untuk mempelajari hubungan temporal. Sistem ini memproses data video dengan membagi video menjadi frame, mengekstrak informasi dari tiap frame, dan mengklasifikasikan konten sebagai kekerasan atau non-kekerasan. Dengan menggunakan dataset kekerasan dunia nyata, sistem mencapai akurasi 91% pada data pengujian, lebih baik dibandingkan penelitian sebelumnya untuk dataset yang sama. Sistem ini menawarkan deteksi cepat dan akurat dengan sumber daya komputasi yang terbatas, namun pengembangan dataset yang lebih besar dan lebih beragam diperlukan untuk meningkatkan generalisasi.

**Tabel 2. 1 Tinjauan Studi**

| No | Peneliti                      | Judul                                                                             | Sumber                                                                                                                                   | Kontribusi                                                                                                                     |
|----|-------------------------------|-----------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------|
| 1  | (Perseghin and Foresti, 2023) | A Shallow System Prototype for Violent Action Detection in Italian Public Schools | Information<br>2023 / No<br>14 / Vol 246<br>/<br><a href="https://doi.org/10.3390/info14040240">https://doi.org/10.3390/info14040240</a> | CNN dapat mendeteksi tindakan kekerasan di lingkungan sekolah dengan perangkat keras berbiaya rendah dan mencapai akurasi 95%. |
| 2  | (G. and Saleem, 2024)         | Violence detection in crowd videos using nuanced                                  | Systems and Soft                                                                                                                         | Mengembangkan metode pendeteksian kekerasan dalam video kerumunan                                                              |

| No | Peneliti                     | Judul                                                                                 | Sumber                                                                                                                                                                                                                                                                                 | Kontribusi                                                                                                                                                                                                   |
|----|------------------------------|---------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|    |                              | facial expression analysis                                                            | Computing / Vol 6 / <a href="https://doi.org/10.1016/j.sasc.2024.200104">https://doi.org/10.1016/j.sasc.2024.200104</a>                                                                                                                                                                | menggunakan CRBM yang mendeteksi dari ekspresi wajah seperti marah, cemas, dan takut dan menghasilkan tingkat akurasi sebesar 88.25%.                                                                        |
| 3  | (Haiura and Iftene, 2024)    | Detecting Violence in Videos using Convolutional Neural Networks                      | Procedia Computer Science / Vol 246 / <a href="https://www.sciencedirect.com/science/article/pii/S1877050924025079?via=ihub">https://www.sciencedirect.com/science/article/pii/S1877050924025079?via=ihub</a>                                                                          | Mengembangkan model deteksi kekerasan berbasis 3D Convolutional Neural Network (CNN) untuk mendeteksi tindakan kekerasan dalam video secara real-time. Menggunakan model v8n mencapai tingkat akurasi 87.25% |
| 4  | (Ullah <i>et al.</i> , 2019) | Violence detection using spatiotemporal features with 3D convolutional neural network | Sensors, 2019 / No 3 / Vol 9 / <a href="https://www.semanticscholar.org/paper/Violence-detection-using-spatiotemporal-features-Ullah/953a96b6cf39d9acab182ea">https://www.semanticscholar.org/paper/Violence-detection-using-spatiotemporal-features-Ullah/953a96b6cf39d9acab182ea</a> | Sistem ini dapat menganalisis fitur spasial dan temporal dari video pengawasan dengan 3 tahap yaitu deteksi orang, ekstraksi fitur dan klasifikasi kekerasan. Penelitian mencapai tingkat akurasi diatas 96% |



| No | Peneliti                                | Judul                                                                                  | Sumber                                                                                                                                                         | Kontribusi                                                                                                                                                                                                                                                                            |
|----|-----------------------------------------|----------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|    |                                         |                                                                                        | <a href="#">6345d0202210ebcc0</a>                                                                                                                              |                                                                                                                                                                                                                                                                                       |
| 5  | (Vosta and Yow, 2022)                   | A CNN-RNN Combined Structure for Real-World Violence Detection in Surveillance Cameras | <i>Appl. Sci.</i> 2022 / No 1 / Vol 3 / <a href="https://doi.org/10.3390/app12031021">https://doi.org/10.3390/app12031021</a> .                                | Sistem ini menggunakan kombinasi arsitektur ResNet50 dan ConvLSTM untuk mendeteksi perilaku abnormal pada data video pengawasan. Sistem ini berhasil mencapai AUC sebesar 81,71%                                                                                                      |
| 6  | (Vieira <i>et al.</i> , 2022)           | Low-Cost CNN for Automatic Violence Recognition on Embedded System                     | Digital Object Identifier / Vol 10 / <a href="https://ieeexplore.ieee.org/abstract/document/9722878">https://ieeexplore.ieee.org/abstract/document/9722878</a> | mengembangkan sistem deteksi kekerasan berbasis arsitektur Mobile Convolutional Neural Networks (CNNs) yang diadaptasi untuk pengenalan otomatis tindakan kekerasan dalam video sehingga memungkinkan untuk menerapkan dengan biaya rendah. Sistem ini mencapai akurasi hingga 92,05% |
| 7  | (Vijeikis, Raudonis and Dervinis, 2022) | Efficient Violence Detection in Surveillance                                           | Sensors 2022 / Vol 22 / <a href="https://doi.org/10.3390/s22062216">https://doi.org/10.3390/s22062216</a>                                                      | Mengembangkan sistem deteksi kekerasan berbasis video pengawasan menggunakan arsitektur U-Net yang memanfaatkan MobileNet V2 sebagai encoder untuk ekstraksi fitur spasial. Hasil eksperimen menunjukkan akurasi rata-rata 82% dengan                                                 |

| No | Peneliti                                       | Judul                                                                                                                                                   | Sumber                                                                                                                                                                                                         | Kontribusi                                                                                                                                                                                                                                                                                                                               |
|----|------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|    |                                                |                                                                                                                                                         |                                                                                                                                                                                                                | presisi 81% pada dataset kompleks                                                                                                                                                                                                                                                                                                        |
| 8  | (Aghassi Zulfikar, Somantri and Sudjadi, 2021) | Penerapan Metode Convolutional Neural Network (CNN) Dan Long Short Term Memory (LSTM) Untuk Pengenalan Aktivitas Manusia Pada CCTV Di Area Tambak Udang | <i>urnal Ilmiah Teknik Elektro / No 9 / Vol 5 /</i><br><a href="https://ejournal3.undip.ac.id/index.php/transient/article/view/29530">https://ejournal3.undip.ac.id/index.php/transient/article/view/29530</a> | Sistem ini mendeteksi dan mengenali aktivitas manusia, seperti berjalan dan berlari, melalui rekaman CCTV dengan tahapan deteksi objek menggunakan YOLO, pelacakan dengan DeepSort, dan pengenalan aktivitas menggunakan kombinasi CNN dan LSTM. Hasil pengujian menunjukkan akurasi rata-rata 79,25%                                    |
| 9  | (NURYASIN, MACHBUB and YULIANTI, 2023)         | Kombinasi Deteksi Objek, Pengenalan Wajah dan Perilaku Anomali menggunakan State Machine untuk Kamera Pengawas                                          | <i>Elkomika / No 1 / Vol 11 /</i><br><a href="https://ejurnal.itenas.ac.id/index.php/elkomika/article/view/7647">https://ejurnal.itenas.ac.id/index.php/elkomika/article/view/7647</a>                         | Menggunakan 3 modul yaitu deteksi objek menggunakan HOG-SVM, pengenalan wajah dengan CNN berbasis arsitektur VGG-16, dan deteksi perilaku anomali dengan spatiotemporal autoencoder berbasis threshold. Hasil pengujian menunjukkan akurasi 88% untuk deteksi objek, 98% untuk pengenalan wajah, dan 78% untuk deteksi perilaku anomali. |
| 10 | (Shaheer and U, 2023)                          | Real-Time Video Violence Detection Using CNN                                                                                                            | <i>International Journal for Research in Applied Science &amp;</i>                                                                                                                                             | Dapat mendeteksi kekerasan berbasis video secara real-time menggunakan arsitektur MobileNet V2 sebagai                                                                                                                                                                                                                                   |

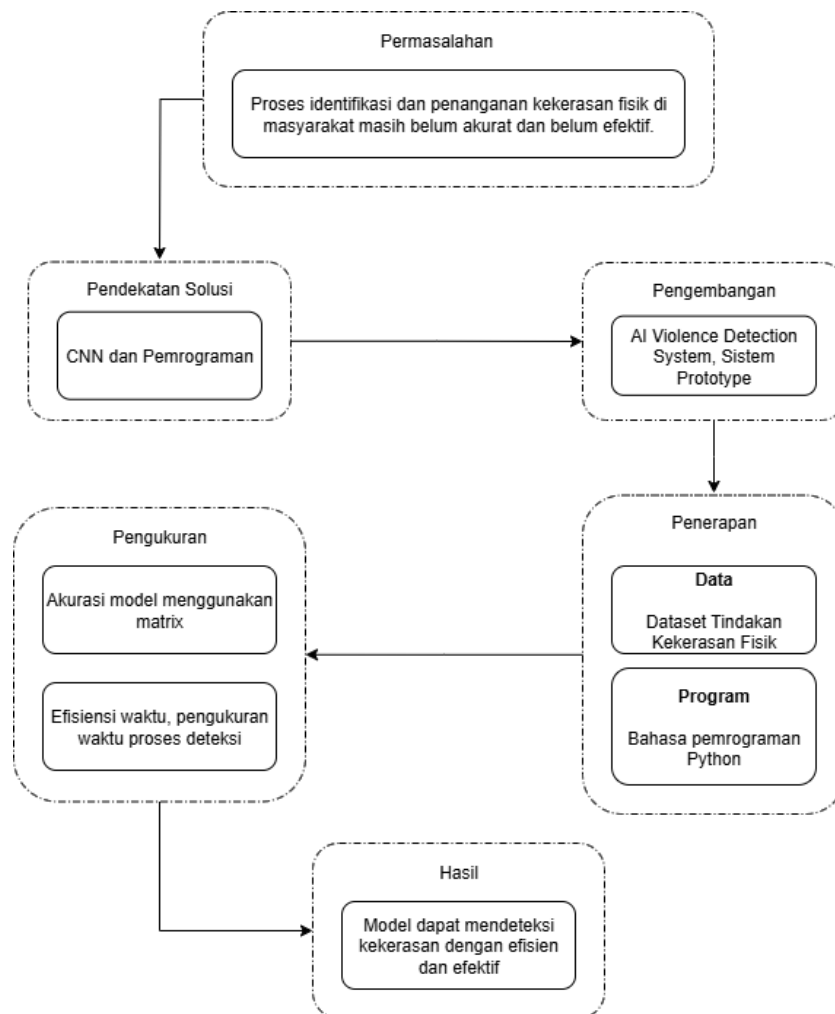
| No | Peneliti | Judul | Sumber                                                                                                                                                                                                                                  | Kontribusi                                                                                                             |
|----|----------|-------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------|
|    |          |       | Engineering<br>Technology<br>(IJRASET)<br>/ Vol 11 /<br><a href="https://www.ijraset.com/best-journal/realtime-video-violence-detection-using-cnn">https://www.ijraset.com/best-journal/realtime-video-violence-detection-using-cnn</a> | ekstraktor fitur spasial dan LSTM untuk mempelajari hubungan temporal. Sistem mencapai akurasi 91% pada data pengujian |

Kontribusi baru dari penelitian ini terletak pada pengembangan sistem deteksi kekerasan fisik berbasis CNN yang dikombinasikan dengan LSTM untuk memperkuat kemampuan analisis spasial dan temporal pada rekaman video. Tidak hanya berfokus pada satu jenis input, sistem ini dirancang untuk mendeteksi kekerasan baik dari video unggahan maupun dari umpan kamera secara real-time. Faktor pembeda lainnya adalah adanya integrasi sistem dokumentasi otomatis, di mana setiap kejadian kekerasan yang terdeteksi akan menghasilkan penyimpanan frame beserta *timestamp* ke dalam basis data terstruktur. Dokumentasi ini memudahkan tindak lanjut oleh pihak berwenang, sehingga sistem tidak hanya bertindak sebagai alat pendeteksi, tetapi juga sebagai instrumen pendukung pelaporan dan evaluasi kasus.

Selain itu, penelitian ini mencoba menghadirkan antarmuka berbasis web yang interaktif dan dapat digunakan diberbagai perangkat. Dengan pendekatan ini, sistem dapat dimanfaatkan secara luas di berbagai lingkungan, baik rumah tangga, ruang publik, maupun area komersial, tanpa membutuhkan perangkat keras khusus atau sumber daya komputasi tinggi. Inovasi ini diharapkan memberikan kontribusi praktis dalam upaya pencegahan kekerasan fisik melalui teknologi yang adaptif dan ramah pengguna.

### C. Kerangka Berpikir

Kerangka berpikir sebagaimana gambar 2.3 dibawah ini:



**Gambar 2. 14 Kerangka Berfikir**

Gambar 2.3 menjelaskan kerangka pemikiran mulai dari permasalahan yang ada sampai kepada hasil yang diperoleh. Penjelasan lebih detail mengenai kerangka pemikiran pada penelitian ini adalah:

- (1) Identifikasi permasalahan tersebut menunjukkan bahwa proses identifikasi dan penanganan kekerasan fisik di masyarakat masih dilakukan secara konvensional, sehingga sering kali tidak akurat, tidak konsisten, dan sarat subjektivitas;
- (2) Diterapkannya pemodelan CNN untuk mendeteksi kekerasan fisik di masyarakat, Model ini dilatih dengan dataset video rekaman untuk mendeteksi dan mengenali gerakan-gerakan yang merupakan suatu tindakan kekerasan fisik;
- (3) Sistem yang akan dikembangkan yaitu AI Violation Detection System yang merupakan sistem yang bisa mendeteksi kekerasan fisik di masyarakat, berbasis prototype sistem website yang akan dihasilkan dalam penelitian ini;
- (4) Data yang akan digunakan untuk melatih model CNN berupa dataset video rekaman dengan indikasi kekerasan fisik dan serta bukan kekerasan, serta bahasa

pemrograman yang akan digunakan adalah python dengan beberapa pustaka seperti TensorFlow atau PyTorch dalam implementasi CNN;

- (5) Untuk metode yang digunakan dalam mengukur akurasi model akan menggunakan matrix seperti precision, Recall, F1-Score dan Confusion Matrix untuk bisa mengevaluasi performa dari model, juga akan diukur efisiensi dari sistem atau aplikasi dengan melibatkan ahli dan pengguna dengan mengacu pada pertanyaan pada PSSUQ yang dimana akan mengacu pada skala likert dan guttman, ini dilakukan untuk dilihat efisiensi dan keakuratan dari sistem apakah sudah mencapai hasil yang diinginkan atau tidak;
- (6) Diterapkannya sistem deteksi berbasis CNN, pengawasan terhadap tindakan kekerasan fisik di masyarakat dapat dilakukan dengan lebih efektif, akurat dan objektif, dibandingkan metode konvensional sebelumnya, dengan diterapkannya model CNN diharapkan mampu mencapai akurasi tinggi dalam hal deteksi kekerasan fisik di masyarakat.

#### **D. Hipotesis Penelitian**

CNN merupakan salah satu algoritma deep learning yang memiliki kemampuan dalam hal klasifikasi. Dalam penelitian – penelitian sebelumnya rata – rata akurasi yang berhasil didapatkan mencapai tingkat 90%. Sistem berbasis pengolahan citra seperti CNN telah digunakan dalam berbagai bidang untuk bisa mengidentifikasi pola secara otomatis dengan tingkat akurasi yang tinggi. Dengan kemampuan tersebut CNN berpotensi diterapkan dalam sistem untuk mendeteksi kekerasan fisik di masyarakat. Dengan pemahaman teoritis dan pemahaman lain diatas maka hipotesis penelitian & pengembangan ini Pemodelan Machine Learning untuk deteksi Kekerasan Fisik berbasis Convolutional Neural Network (CNN) diduga akurat dan efektif untuk mendeteksi kekerasan fisik di masyarakat.