

## IMPLEMENTASI METODE LEVENSTHEIN DISTANCE DAN COSINE SIMILARITY UNTUK DETEKSI KEMIRIPAN GAMBAR

Niko Surya Atmaja

Sistem Komputer, Universitas Pembangunan Panca Budi, Indonesia  
Email: [niko.suryaatmaja@gmail.com](mailto:niko.suryaatmaja@gmail.com)

**Abstrak:** Gambar adalah media visual dua dimensi di atas bidang yang tidak transparan. Sebuah gambar dapat di edit dan meniru bentuk gambar yang lain dengan menggunakan aplikasi yang tersedia secara gratis. Sehingga dapat terjadi peniruan gambar dengan aplikasi edit gambar, oleh karena itu seseorang yang mencurigai sebuah gambar hasil edit harus membandingkan dengan gambar yang asli. Masalah yang terjadi adalah seseorang yang mencari kemiripan gambar tidak dapat menghasilkan persentase yang tepat. Oleh karena itu dibutuhkan sebuah cara yang dapat mendeteksi dan menghasilkan persentase kemiripan gambar yang tepat. Penelitian ini menggunakan metode *levenshtein distance* untuk mendeteksi dua buah gambar sehingga mendapatkan hasil persentase kemiripan pada dua buah gambar. Akan tetapi penelitian ini juga menggunakan metode *cosine similarity* yang akan digunakan untuk mengetahui persentase kemiripan gambar sehingga dapat mengetahui selisih kemiripan dari penggunaan dua metode pendeteksi kemiripan gambar. Dengan adanya penerapan metode *levenshtein distance* dan *cosine similarity* maka mendeteksi kemiripan gambar dapat dilakukan dengan mudah.

**Kata Kunci:** *Deteksi, Cosine Similarity, Gambar, Implementasi, Levenshtein Distance.*

**Abstract:** *An image is a two-dimensional visual medium above a non-transparent plane. An image can be edited and mimick the shape of another image by using an app that is available for free. So that there can be an image impersonation with an image editing application, therefore someone who suspects an edited image must compare with the original image. The problem that occurs is that someone looking for a similarity of an image cannot produce the right percentage. Therefore, a way is needed that can detect and produce the right percentage of image similarity. This study used the levenshtein distance method to detect two images so as to get the result of a percentage of similarities in two images. However, this study also uses the cosine similarity method that will be used to determine the percentage of image similarity so that it can find out the difference in similarity from the use of two methods of detecting image likeness. With the application of the levenshtein distance and cosine similarity methods, detecting image similarity can be done easily.*

**Keywords:** *Cosine Similarity, Detection, Image, Implementation, Levenshtein Distance.*

## 1. PENDAHULUAN

Gambar adalah media visual dua dimensi di atas bidang yang tidak transparan. [1]. Sebuah gambar dapat di edit dan meniru bentuk gambar yang lain dengan menggunakan aplikasi yang tersedia secara gratis. Sehingga dapat terjadi peniruan gambar dengan aplikasi edit gambar, oleh karena itu seseorang yang mencurigai sebuah gambar hasil edit harus membandingkan dengan gambar yang asli.[2] Masalah yang terjadi adalah seseorang yang mencari kemiripan gambar tidak dapat menghasilkan persentase yang tepat. Oleh karena itu dibutuhkan sebuah cara yang dapat mendeteksi dan menghasilkan persentase kemiripan gambar yang tepat. Penelitian ini mengimplementasikan metode *levensthein distance* untuk mendeteksi dua buah gambar sehingga mendapatkan hasil persentase kemiripan pada dua buah gambar. Implementasi adalah tindakan berdasarkan kemampuan pada suatu penguasaan materi ke dalam situasi yang nyata. [3]. Akan tetapi penelitian ini juga menggunakan metode *cosine similarity* yang akan digunakan untuk mengetahui persentase kemiripan gambar sehingga dapat mengetahui selisih kemiripan dari penggunaan dua metode pendeteksi kemiripan gambar. Deteksi adalah sebuah tindakan untuk menemukan suatu kenyataan, keberadaan dan anggapan. [4]. Kemiripan adalah hasil dari dua objek yang memiliki beberapa nilai yang sama. [5]. Dengan adanya penerapan metode *levensthein distance* dan *cosine similarity* maka mendeteksi kemiripan gambar dapat dilakukan dengan mudah.

*Levensthein Distance* adalah sebuah metode yang dapat digunakan untuk menguji kemiripan dua urutan data. [6]. Rumus dari metode *levensthein distance* disajikan pada persamaan 1.

$$\text{Dist}(A, B) = 1 - \frac{\text{dist}(A,B)}{\max(|A|,|B|)} \quad (1)$$

Dimana:

Dist (A,B)= Kemiripan

A = nilai besaran

B = nilai besaran

|A| = determinan nilai besaran A

|B| = determinan nilai besaran B

Pada metode *Levenshtein distance*, terdapat beberapa langkah sebagai berikut:

- a. Operasi Pengubah.
- b. Operasi Penambahan.
- c. Operasi Penghapusan. [7].

*Cosine Similarity* terdiri dari dua kata yaitu *Cosine* yang merupakan sudut antara dua besaran yang memiliki nilai berdimensi dan *Similarity* yang merupakan kemiripan. [8]. *Cosine similarity* digunakan untuk mencari satu nilai diantara dua nilai besaran [9]. Sehingga pengertian *Cosine Similarity* yaitu kemiripan dari dua besaran yang memiliki nilai berdimensi. [10]. Rumus dari metode *cosine similarity* disajikan pada persamaan 2.

$$\text{Sim}(A, B) = \frac{A \cdot B}{||A|| \cdot ||B||} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \sqrt{\sum_{i=1}^n (B_i)^2}} \quad (2)$$

Dimana:

Sim(A,B) = Kemiripan

A = nilai besaran

B = nilai besaran

||A|| = determinan nilai besaran

||B|| = determinan nilai besaran

A<sub>i</sub> = bobot dari nilai besaran

B<sub>i</sub> = bobot dari nilai besaran

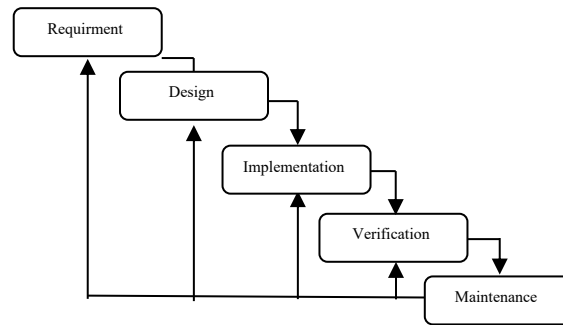
i = perulangan

n = jumlah nilai besaran

∑ (Sigma) = Perulangan rumus. [11].

## 2. METODE PENELITIAN

Metode untuk melaksanakan penelitian dilakukan dengan beberapa tahapan diantaranya *requirement*, *design*, *implementation*, *verification*, dan *maintenance*. Tahapan metode untuk melaksanakan penelitian disajikan pada gambar 1.



Gambar 1. Diagram Alir Penelitian

Penjelasan :

- a. *Requirement*  
Seluruh kebutuhan untuk menyelesaikan penelitian ini adalah gambar dengan berbagai format.
- b. *Design*  
*Design* pada sistem menggunakan pemrograman *microsoft visual basic 2010*.
- c. *Implementation*  
Metode *levensthein distance* dan *cosine similarity* diterapkan pada pemrograman *microsoft visual basic 2010*.
- d. *Verification*  
Pengujian dari penerapan metode yang digunakan untuk mendeteksi kemiripan gambar.
- e. *Maintenance*  
Setiap hasil uji dari aplikasi yang memiliki hasil yang tidak baik maka dilakukan perbaikan dan mengulang pada tahapan metode penelitian yang dibutuhkan

## 3. HASIL DAN PEMBAHASAN


### 3.1 Metode *Levensthein Distance* dan *Cosine Similarity*

Metode *levensthein distance* dan *cosine similarity* digunakan untuk memperoleh persentase kemiripan dari dua buah gambar.

Contoh:

Terdapat dua buah gambar yang tampak sama untuk dideteksi menggunakan metode *levensthein distance* yang disajikan pada Tabel 1.

Tabel 1. Gambar Yang Akan Dideteksi

No.	Nama Gambar	Gambar
1.	Mobil.jpg	

2. Mobil.jpg



### 3.2 Metode Levensthein Distance

Levensthein Distance adalah sebuah metode yang dapat digunakan untuk menguji kemiripan dua urutan data. [6]. Rumus dari metode *levensthein distance* disajikan pada persamaan 1.

$$\text{Dist}(A, B) = 1 - \frac{\text{dist}(A,B)}{\max(|A|,|B|)} \quad (1)$$

Dimana:

Dist (A,B)= Kemiripan

A = nilai besaran

B = nilai besaran

|A| = determinan nilai besaran A

|B| = determinan nilai besaran B

Pada metode *Levenshtein distance*, terdapat beberapa langkah sebagai berikut:

- a. Operasi Pengubah.
- b. Operasi Penambahan.
- c. Operasi Penghapusan. [7].

Solusi:

- a. Operasi Pengubah.

Tindakan untuk mengubah gambar menjadi sebuah nilai. Pada sebuah gambar terdapat nilai *Red, Green* dan *Blue* (RGB) yang berasal dari nilai citra gambar. Sehingga nilai RGB digunakan untuk mendeteksi kemiripan dari dua gambar dan disajikan pada Tabel 2.

**Tabel 2.** Nilai RGB Gambar

No.	Gambar	RGB		
1.		R:15	R:100	R:180
		G:13	G:57	G:48
		B:12	B:47	B:45
		R:200	R:205	R:214
		G:25	G:20	G:27
		B:37	B:12	B:28
		R:211	R:200	R:102
		G:17	G:13	G:37
		B:23	B:15	B:48
2.		R:15	R:100	R:180
		G:13	G:57	G:48
		B:12	B:47	B:45
		R:200	R:205	R:214
		G:26	G:20	G:26
		B:37	B:16	B:22
		R:211	R:200	R:102
		G:17	G:14	G:39
		B:23	B:15	B:49

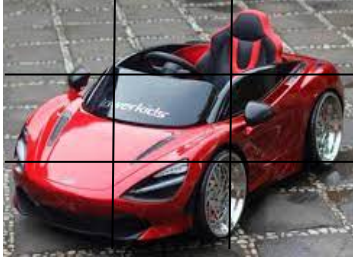
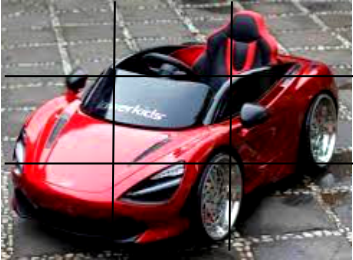
- b. Operasi Penambahan.

Tindakan untuk melakukan penambahan pixel jika nilai pixelnya berbeda. Pada contoh yang digunakan, pixel bernilai sama yaitu 3x3 sehingga tidak terjadi penambahan nilai pixel untuk penilaian kemiripan.

c. Operasi Penghapusan.

Tindakan untuk menghapus beberapa nilai yang sama dan disajikan pada Tabel 3.

**Tabel 3.** Nilai RGB Gambar Operasi Penghapusan

No.	Gambar	RGB		
1.				
		G:25		G:27
		R:211	R:200 G:13	R:102 G:37
2.				
		G:26		G:26
		R:211	R:200 G:14	R:102 G:39

Sehingga dapat dihitung kemiripan dua buah gambar dengan rumus metode *levensthein distance* sebagai berikut:

R: 6 nilai yang sama

G: 5 nilai yang sama

B: 9 nilai yang sama

$$Dist(R) = 1 - \frac{dist(R)}{\max(|R|)}$$

$$Dist(A, B) = 1 - \frac{6}{9}$$

$$Dist(A, B) = 1 - 0.6667$$

$$Dist(A, B) = 0.3$$

$$Dist(G) = 1 - \frac{dist(G)}{\max(|G|)}$$

$$Dist(A, B) = 1 - \frac{5}{9}$$

$$Dist(A, B) = 1 - 0.5556$$

$$Dist(A, B) = 0.4$$

$$Dist(B) = 1.0 - \frac{dist(B)}{\max(|B|)}$$

$$Dist(A, B) = 1 - \frac{9}{9}$$

$$Dist(A, B) = 1 - 1$$

$$Dist(A, B) = 0$$

$$Total = Dist(R) + Dist(G) + Dist(B)$$

$$Dist(A, B) = 0.3 + 0.4 + 0$$

$$Dist(A, B) = 0.7$$

Persentase =  $0.7 * 100\%$   
Dist(A, B) = 70% Mirip

### 3.3 Metode Cosine Similarity

Cosine Similarity terdiri dari dua kata yaitu Cosine yang merupakan sudut antara dua besaran yang memiliki nilai berdimensi dan Similarity yang merupakan kemiripan. [8]. Cosine similarity digunakan untuk mencari satu nilai diantara dua nilai besaran [9]. Sehingga pengertian Cosine Similarity yaitu kemiripan dari dua besaran yang memiliki nilai berdimensi. [10]. Rumus dari metode cosine similarity disajikan pada persamaan 2.

$$\text{Sim}(A, B) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \sqrt{\sum_{i=1}^n (B_i)^2}} \quad (2)$$

Dimana:

Sim(A,B) = Kemiripan

A = nilai besaran

B = nilai besaran

$\|A\|$  = determinan nilai besaran

$\|B\|$  = determinan nilai besaran

$A_i$  = bobot dari nilai besaran

$B_i$  = bobot dari nilai besaran

i = perulangan

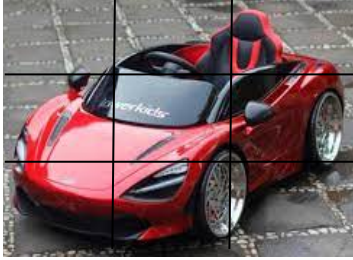
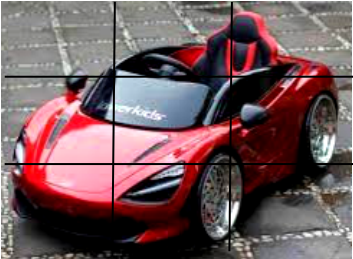
n = jumlah nilai besaran

$\Sigma$  (Sigma) = Perulangan rumus. [11].

Solusi:

Penerapan cosine similarity untuk perhitungan deteksi kemiripan gambar disajikan pada Tabel 4.

**Tabel 4.** Nilai RGB Gambar

No.	Gambar	RGB		
1.		R:15 G:13 B:12 R:200 G:25 B:37 R:211 G:17 B:23	R:100 G:57 B:47 R:205 G:20 B:12 R:200 G:13 B:15	R:180 G:48 B:45 R:214 G:27 B:28 R:102 G:37 B:48
2.		R:15 G:13 B:12 R:200 G:26 B:37 R:211 G:17 B:23	R:100 G:57 B:47 R:205 G:20 B:16 R:200 G:14 B:15	R:180 G:48 B:45 R:214 G:26 B:22 R:102 G:39 B:49

Dari kedua gambar maka diperoleh data sebagai berikut:

Red:

15:pada gambar pertama ada 1, gambar kedua ada 1.

100:pada gambar pertama ada 1, gambar kedua ada 1.

180:pada gambar pertama ada 1, gambar kedua ada 1.

200:pada gambar pertama ada 1, gambar kedua ada 1.

205:pada gambar pertama ada 1, gambar kedua ada 1.

214:pada gambar pertama ada 1, gambar kedua ada 1.

211:pada gambar pertama ada 1, gambar kedua ada 1.

200:pada gambar pertama ada 1, gambar kedua ada 1.

102:pada gambar pertama ada 1, gambar kedua ada 1.

Green:

13:pada gambar pertama ada 1, gambar kedua ada 1.

57:pada gambar pertama ada 1, gambar kedua ada 1.

48:pada gambar pertama ada 1, gambar kedua ada 1.

25:pada gambar pertama ada 1, gambar kedua ada 0.

20:pada gambar pertama ada 1, gambar kedua ada 1.

27:pada gambar pertama ada 1, gambar kedua ada 0.

17:pada gambar pertama ada 1, gambar kedua ada 1.

13:pada gambar pertama ada 1, gambar kedua ada 0.

37:pada gambar pertama ada 1, gambar kedua ada 0.

Blue:

12:pada gambar pertama ada 1, gambar kedua ada 1.

47:pada gambar pertama ada 1, gambar kedua ada 1.

45:pada gambar pertama ada 1, gambar kedua ada 1.

37:pada gambar pertama ada 1, gambar kedua ada 1.

12:pada gambar pertama ada 1, gambar kedua ada 0.

28:pada gambar pertama ada 1, gambar kedua ada 0.

23:pada gambar pertama ada 1, gambar kedua ada 1.

15:pada gambar pertama ada 1, gambar kedua ada 1.

48:pada gambar pertama ada 1, gambar kedua ada 0.

$$\text{Similarity (A,B)} = \frac{A \cdot B}{\|A\| \cdot \|B\|} = \frac{\sum_{i=1}^n A \cdot B}{\sqrt{\sum_{i=1}^n (A_i)^2} \sqrt{\sum_{i=1}^n (B_i)^2}}$$

$$\begin{aligned} \text{Similarity (R)} &= \frac{(1 * 1) + (1 * 1) + (1 * 1) + (1 * 1) + (1 * 1) + (1 * 1) + (1 * 1) + (1 * 1) + (1 * 1)}{\sqrt{1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2} * \sqrt{1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2}} \\ &= \frac{9}{\sqrt{9} * \sqrt{9}} = 1 \end{aligned}$$

$$\text{Similarity (R)} = \frac{9}{\sqrt{9} * \sqrt{9}} = 1$$

$$\begin{aligned} \text{Similarity (G)} &= \frac{(1 * 1) + (1 * 1) + (1 * 1) + (1 * 0) + (1 * 1) + (1 * 0) + (1 * 1) + (1 * 0) + (1 * 0)}{\sqrt{1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2} * \sqrt{1^2 + 1^2 + 1^2 + 0^2 + 1^2 + 0^2 + 1^2 + 0^2 + 0^2}} \\ &= \frac{5}{\sqrt{9} * \sqrt{5}} = 0.75 \end{aligned}$$

$$\text{Similarity (G)} = \frac{5}{\sqrt{9} * \sqrt{5}} = 0.75$$

$$\begin{aligned} \text{Similarity (B)} &= \frac{(1 * 1) + (1 * 1) + (1 * 1) + (1 * 1) + (1 * 0) + (1 * 0) + (1 * 1) + (1 * 1) + (1 * 0)}{\sqrt{1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2} * \sqrt{1^2 + 1^2 + 1^2 + 1^2 + 0^2 + 0^2 + 1^2 + 1^2 + 0^2}} \\ &= \frac{6}{\sqrt{9} * \sqrt{6}} = 0.73 \end{aligned}$$

$$\text{Similarity (R)} = \frac{6}{\sqrt{9} * \sqrt{6}} = 0.73$$

$$\text{Total} = \frac{\text{Similarity(R)} + \text{Similarity(G)} + \text{Similarity(B)}}{3}$$

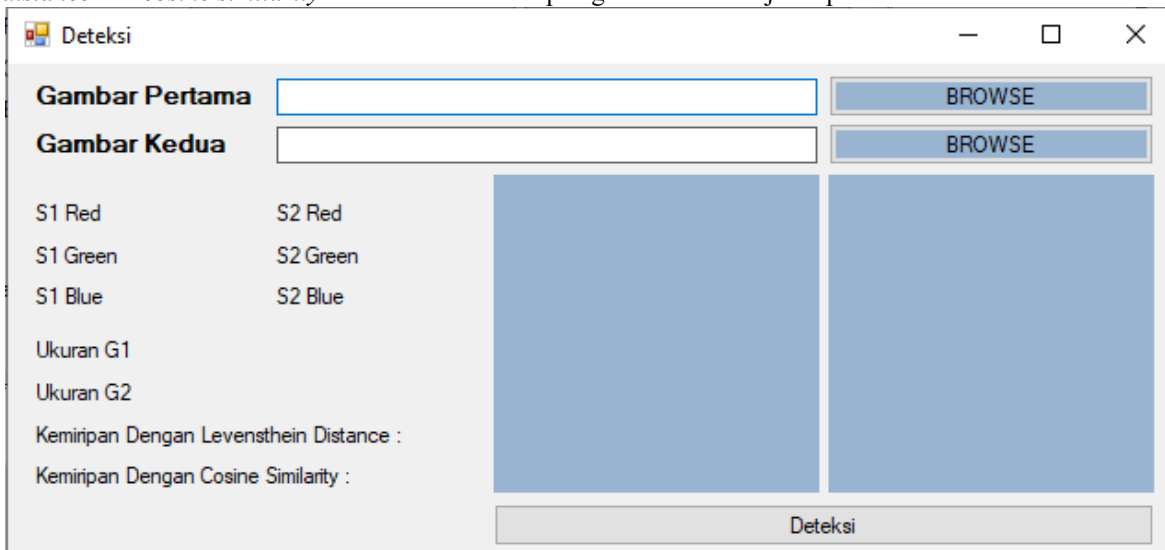
$$\text{Dist(A, B)} = \frac{1 + 0.75 + 0.73}{3}$$

$$\text{Dist(A, B)} = 0.83$$

$$\text{Persentase} = 0.83 * 100\%$$

$$\text{Dist(A, B)} = 83\% \text{ Mirip}$$

Hasil yang diperoleh pada penelitian ini yaitu sebuah aplikasi yang mengimplementasikan metode *levensthein distance* dan *cosine similarity* untuk deteksi kemiripan gambar dan disajikan pada Gambar 1.



**Gambar 1.** Aplikasi Deteksi Kemiripan Gambar

**Tabel 5.** Pengujian Aplikasi Menggunakan Black Box Testing

No	Form Pengecekan	Keterangan	Hasil
1.	Klik <i>Button Browse</i> Pada Gambar Pertama	Sistem akan menampilkan jendela pemilihan <i>file</i>	Sesuai
2.	Klik <i>Button Browse</i> Pada Gambar Kedua	Sistem akan menampilkan jendela pemilihan <i>file</i>	Sesuai
3.	Klik <i>Button Deteksi</i>	Sistem akan menampilkan hasil deteksi dari metode <i>levensthein distance</i> dan <i>cosine similarity</i>	Sesuai

#### 4. KESIMPULAN DAN SARAN

##### 4.1. Kesimpulan

Dari hasil dan pembahasan diperoleh beberapa kesimpulan yaitu:

- Metode *levensthein distance* dapat digunakan untuk deteksi kemiripan gambar dan diperoleh dari hasil pengujian pada dua buah gambar sebesar 70%.
- Metode *cosine similarity* dapat digunakan untuk deteksi kemiripan gambar dan diperoleh dari hasil pengujian pada dua buah gambar sebesar 83%.

##### 4.2. Saran

Dari kekurangan yang terdapat pada penelitian ini maka saran dari penelitian ini yaitu:

- Sebaiknya aplikasi dapat mendeteksi kemiripan lebih dua buah gambar.
- Sebaiknya aplikasi dapat diterapkan secara *online* sehingga dapat digunakan dimana saja.
- Sebaiknya aplikasi menggunakan banyak pilihan metode pendeteksi.



**DAFTAR PUSTAKA**

- [1] N. Hidayati and L. Yuliandani, “Pengaruh Beauty Vlogger, Citra Merek Dan Label Halal Terhadap Minat Beli Produk Kosmetik Wardah,” *JSMA (Jurnal Sains Manajemen dan Akuntansi)*, vol. 12, no. 1, pp. 65–76, 2020, doi: 10.37151/jsma.v12i1.50.
- [2] B. Yanto, B. -, J. -, and B. H. Hayadi, “INDENTIFIKASI POLA AKSARA ARAB MELAYU DENGAN JARINGAN SYARAF TIRUAN CONVOLUTIONAL NEURAL NETWORK (CNN),” *JSAI (Journal Scientific and Applied Informatics)*, vol. 3, no. 3, 2020, doi: 10.36085/jsai.v3i3.1151.
- [3] I. Magdalena, A. Salsabila, A. Krianasari, D., and F. Apsarini, S., “Implementasi Model Pembelajaran Daring Pada Masa Pandemi Covid-19 Di Kelas III SDN Sindangsari III,” *Jurnal Pendidikan dan Dakwah*, vol. 3, no. 1, pp. 119–128, 2020.
- [4] J. M. S. Waworundeng, “Desain Sistem Deteksi Asap dan Api Berbasis Sensor, Mikrokontroler dan IoT,” *CogITo Smart Journal*, vol. 6, no. 1, p. 117, 2020, doi: 10.31154/cogito.v6i1.239.117-127.
- [5] A. Zouhir, R. El Ayachi, and M. Biniz, “A comparative Plagiarism Detection System methods between sentences,” *Journal of Physics: Conference Series*, vol. 1743, no. 1, 2021, doi: 10.1088/1742-6596/1743/1/012041.
- [6] C. Adjetey and K. S. Adu-Manu, “Content-based Image Retrieval using Tesseract OCR Engine and Levenshtein Algorithm,” *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 7, pp. 666–675, 2021, doi: 10.14569/IJACSA.2021.0120776.
- [7] K. Janardhan, “Tendency to Study of the Cross Batch Redundancy Detection of Efficient Image Sharing,” vol. 3, no. 10, p. 3048, 2019.
- [8] F. AZMI, A. Saleh, and N. P. Dharshinni, “Face Identification on Login Security Using Algorithm Combination of Viola-Jones and Cosine Similarity,” *Journal of Informatics and Telecommunication Engineering*, vol. 4, no. 1, pp. 203–211, 2020, doi: 10.31289/jite.v4i1.3885.
- [9] M. S. Stupariu, A. I. Pleșoianu, I. Pătru-Stupariu, and C. Fürst, “A method for tree detection based on similarity with geometric shapes of 3D geospatial data,” *ISPRS International Journal of Geo-Information*, vol. 9, no. 5, 2020, doi: 10.3390/ijgi9050298.
- [10] B. Kim *et al.*, “Structural health monitoring with sensor data and cosine similarity for multi-damages,” *Sensors (Switzerland)*, vol. 19, no. 14, 2019, doi: 10.3390/s19143047.
- [11] S. Saha *et al.*, “Feature selection for facial emotion recognition using cosine similarity-based harmony search algorithm,” *Applied Sciences (Switzerland)*, vol. 10, no. 8, pp. 1–22, 2020, doi: 10.3390/APP10082816.