

# Pengenalan Citra Wajah pada Sistem Presensi Karyawan Menggunakan Metode K-Nearest Neighbor dan Haversine Formula

Muh. Mustafa Daniel Noya<sup>1\*</sup>, Mardi Hardjianto<sup>2</sup>

<sup>1,2</sup>Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur  
Jl. Ciledug Raya, RT.10/RW.2, Petukangan Utara, Pesanggrahan, Jakarta, Indonesia

e-mail koresponden: [tafanoya@gmail.com](mailto:tafanoya@gmail.com)

(received: 28/07/2023, revised: 14/08/2023, accepted: 15/08/2023)

## Abstrak

Penerapan pengolahan citra wajah pada sistem presensi karyawan menjadi kebutuhan penting di era teknologi. PT. SISMEDIKA telah melakukan implementasi sistem presensi karyawan dengan menggunakan teknologi sidik jari (*finger print*). Namun, hal tersebut dapat menghambat karyawan jika terjadi cedera pada jari ataupun jari dalam keadaan basah. Oleh karena itu, penelitian ini membuat aplikasi presensi dengan pendeteksian wajah menggunakan metode *K-Nearest Neighbor*. Sistem ini juga dilengkapi teknologi *geotagging* dengan GPS dan pemanfaatan *Google Maps API* untuk memilih dan membatasi area presensi dan melakukan perhitungan jarak antara kedua titiknya dengan metode *Haversine Formula*. Penelitian ini bertujuan untuk menghindari faktor-faktor yang dapat menjadi penghambat bagi karyawan dalam melakukan presensi. Hasil penelitian menunjukkan metode *Haversine Formula* berjalan baik, karyawan tidak dapat presensi di luar area yang ditentukan. Penggunaan dataset foto individu dengan 10 label. Identifikasi wajah menggunakan K-NN mencapai akurasi 0,99784, presisi 0,991, dan recall 0,979.

**Kata Kunci:** Haversine Formula, K-Nearest Neighbor, Pengenalan Citra Wajah, Sistem Presensi Karyawan.

## FACE IMAGE RECOGNITION IN EMPLOYEE ATTENDANCE SYSTEM USING THE K-NEAREST NEIGHBOR METHOD AND HAVERSINE FORMULA

### Abstract

*The application of facial image processing in employee attendance systems is an essential need in the technological era. PT SISMEDIKA has implemented an employee attendance system using fingerprint technology. However, this can hinder employees if there is an injury to the finger or the finger is wet. Therefore, this research uses the K-Nearest Neighbor method to make a presence application with face detection. This research makes a system that features geotagging technology with GPS utilizing the Google Maps API to select and limit the attendance area and calculate the distance between the two points using the Haversine Formula method. This research aims to avoid factors that can hinder employees' attendance. The results show that the Haversine Formula method works well; employees cannot take attendance outside the designated area by using individual photo datasets with ten labels. Face identification using K-NN achieved an accuracy of 0.99784, precision of 0.991, and recall of 0.979.*

**Keywords:** Employee Attendance System, Face Image Recognition, Haversine Formula, K-Nearest Neighbor.

### 1. Pendahuluan

Sistem presensi merupakan aspek yang signifikan dalam kehidupan sehari-hari dimana sistem ini berperan sebagai salah satu metode pencatatan kehadiran seseorang, baik itu di tempat kerja, sekolah, maupun kampus [1]. Penulis lain juga mendefinisikan "Sistem Presensi" sebagai suatu sistem yang bertujuan untuk mengelola kehadiran karyawan dalam sebuah lembaga atau institusi. Sistem ini secara

otomatis mencatat kehadiran karyawan dan menggunakan data kehadiran tersebut sebagai sumber laporan untuk memenuhi kebutuhan pengelolaan karyawan [2]. Pada umumnya sebuah sistem presensi dapat mencatat waktu kehadiran, waktu kepulangan, jumlah hadir, jumlah izin disertai keterangan izin karena sakit atau kepentingan lainnya. Pada saat ini banyak perusahaan yang telah melengkapi kantornya dengan sistem presensi yang saling terintegrasi antara karyawan dan atasan dengan menerapkan teknologi *fingerprnt* untuk sistemnya. Adapun kelemahan dari sistem presensi berbasis *fingerprnt* ini yaitu sensor *finger* yang tidak dapat mendeteksi dengan baik pada kondisi-kondisi tertentu seperti ketika kondisi jari sedang basah, lecet karena luka dan hal lainnya. Hal ini tentu dapat mempersulit karyawan dalam melakukan presensi.

PT. Sistem Integrasi Medika (SISMEDIKA) adalah perusahaan teknologi informasi dengan fokus pengembangan terhadap sistem informasi rumah sakit, klinik dan laboratorium di berbagai kota yang tersebar di Indonesia agar dapat beroperasi secara lengkap dan terintegrasi. Pada saat ini, perusahaan tersebut sedang menjalani proyek pengerjaan sistem rumah sakit di RSPAD Gatot Subroto, Jakarta Pusat. Permasalahan terdapat pada perusahaan yang saat ini masih melakukan presensi dengan sistem dan teknologi *fingerprnt* yang memiliki kelemahan-kelemahan yang telah disebutkan sebelumnya. Oleh karena itu, peneliti bermaksud untuk membuat sistem presensi karyawan berbasis web dengan menerapkan teknik pengenalan wajah pada setiap karyawan sebagai solusi dari permasalahan yang ada.

Program mengidentifikasi karyawan menggunakan pengenalan wajah, yaitu membandingkan pola gambar wajah karyawan dengan pola gambar foto wajah karyawan yang tersimpan di komputer. Pada penelitian ini pengenalan wajah diimplementasikan menggunakan metode *K-Nearest Neighbor* (KNN) dan perhitungan jarak antara dua titik dengan metode *Haversine Formula*. *K-Nearest Neighbor* (k-NN atau KNN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran (*neighbor*) yang jaraknya paling dekat dengan objek tersebut [3].

KNN akan mengelompokkan hasil perhitungan dengan data latih yang mempunyai kerabat terbanyak dalam nilai jangkauan yang ditentukan. Jarak antara data latih dan data uji dihitung menggunakan persamaan Euclidean [4]. Hasil dari penelitian terdahulu dengan judul “Implementasi *Face Recognition* pada Absensi Kehadiran Mahasiswa Menggunakan Metode *Haar Cascade Classifier*” [5] melakukan tahapan dan pengujian penelitian dengan metode *haar cascade* dalam proses pengenalan wajah yaitu sebanyak 25 data wajah dengan ketentuan jarak terdekat 30 cm dan jarak terjauh 150cm dan dihasilkan bahwa pengujian dengan status berhasil dikenali dengan benar sebanyak 19 orang dan status tidak berhasil dikenali dengan benar sebanyak 6 orang. Selanjutnya diperoleh hasil nilai akurasi sebesar 76%. Nilai akurasi tersebut dibandingkan dengan metode K-NN pada proses yang sama yaitu pengenalan wajah sebanyak 10 data wajah karyawan yang berbeda dengan jarak rata-rata 60cm mendapatkan hasil akurasi keseluruhan sebesar 99%.

## 2. Metode Penelitian

### 2.1 Data Penelitian

Data yang digunakan dalam penelitian ini berasal dari lembaga tempat riset ini dilakukan untuk program ini. Data tersebut merupakan informasi mengenai karyawan yang bekerja sebagai tenaga ahli.

Tabel 1. Data Karyawan

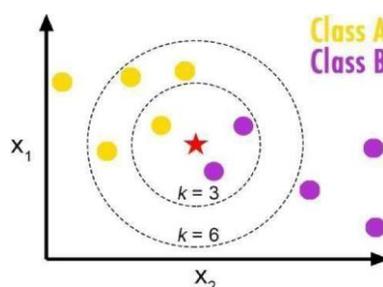
Nama Pegawai	Alamat	No. Telp
Niken	Jl Acordion B1 Q/12, Dki Jakarta	0854-555-507
Rico	Jl Pertanian Raya Komplek SLB A Pembina, Dki Jakarta	0859-555-688
Viko	Jl Tebet Tmr Raya 4, Dki Jakarta	0853-555-793
Laurenza	Jl P Jayakarta 35, Dki Jakarta	0878-555-96
Muthiara	Jl Raya Serdang F/31 RT 007/01, Dki Jakarta	0855-555-513
Riki	Jl Dr Susilo 1, Dki Jakarta	0852-555-120
Novinda	Jl Pinangsia Raya Glodok Harco B1 D/223, Dki Jakarta	0878-555-143
Andreas	Ged Sudirman Tower Kav-60, Jakarta	0878-555-907
Daniel	Jl Falatehan I 34, Jakarta	0838-555-690
Fahri	Jl Jagakarsa I 17, Dki Jakarta	0838-555-190

## 2.2 Penerapan Metode

Metode KNN terdiri dari dua tahap yaitu tahap pembelajaran (*training*) dan tahap klasifikasi atau pengujian (*testing*). Pada tahap *training*, algoritma ini hanya menyimpan vektor-vektor fitur dan klasifikasi dari data *training*. Sedangkan pada tahap klasifikasi, fitur-fitur yang sama dihitung untuk data yang akan diuji (yang belum diklasifikasikan). Jarak antara vektor baru ini dengan seluruh vektor data pembelajaran dihitung dan beberapa  $k$  tetangga terdekat diambil. Perhitungan jarak ketetanggaan menggunakan algoritma *euclidean* sebagaimana ditunjukkan dalam persamaan 1.

$$euc = \sqrt{(a_1 - b_1)^2 + \dots + (a_n - b_n)^2} \quad (1)$$

Dimana  $a = a_1, a_2, \dots, a_n$ , dan  $b = b_1, b_2, \dots, b_n$  mewakili  $n$  nilai atribut dari dua *record* untuk atribut dengan nilai kategori. Sebuah titik akan diprediksi jenisnya berdasarkan pada klasifikasi terbanyak dari *neighbor* di sekitarnya, ilustrasinya dapat dilihat pada Gambar 1 dibawah ini:



**Gambar 1.** Ilustrasi penggunaan nilai  $k$  pada metode KNN

*Haversine formula* adalah metode perhitungan jarak antara dua titik di bumi berdasarkan panjang garis lurus antara keduanya, dengan memperhitungkan kelengkungan bumi [5]. Formula *Haversine* terdapat pada persamaan 2.

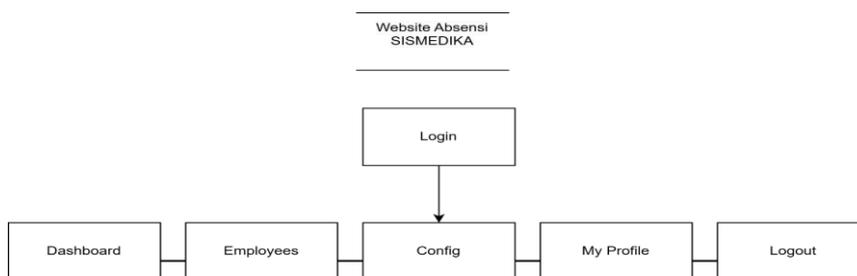
$$\begin{aligned} x &= (\text{lon}2 - \text{lon}1) \cdot \cos\left(\frac{\text{lat}1 + \text{lat}2}{2}\right) \\ y &= (\text{lat}2 - \text{lat}1) \\ d &= \sqrt{(x \cdot x) + (y \cdot y)} \cdot R \end{aligned} \quad (2)$$

### Keterangan:

$lat1$	= latitude awal
$lon$	= longitude awal
$lat2$	= latitude tujuan
$lon2$	= longitude tujuan
$x$	= longitude (lintang)
$y$	= latitude (bujur)
$d$	= jarak
$R$	= jari-jari bumi sebesar 6371 (km) 1 derajat = 0,0174532925

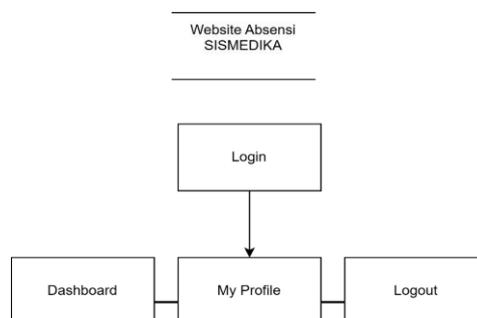
## 2.3 Rancangan Menu

Struktur menu di bawah ini merupakan rancangan menu website yang terbagi menjadi dua yaitu tampilan menu dari sisi *manajer* dan *employee*. Pada user *manager* berikut terdapat lima menu yang dapat digunakan. Setelah user login dengan status sebagai *manager* maka user akan mengakses tampilan menu *dashboard* sebagai menu *default*. Selanjutnya user dapat mengakses menu *employees* untuk melihat dan menambahkan user akun karyawan, menu *config*, *my profile* dan menu *logout* untuk keluar dari aplikasi.



Gambar 2. Rancangan Menu *Manager*

Selanjutnya, pada rancangan menu user karyawan terdapat tiga menu yang dapat digunakan. Setelah user *login* dengan status sebagai *employee* maka user akan mengakses tampilan menu *dashboard* sebagai menu *default*. User karyawan dapat mengakses menu *my profile* dan menu *logout* untuk keluar dari aplikasi.



Gambar 3. Rancangan Menu *Employee*

### 3 Hasil Dan Pembahasan

Beberapa tahapan metode yang digunakan dalam penelitian ini yaitu melakukan penentuan dan pengambilan data titik lokasi karyawan untuk melakukan presensi. Jika titik lokasi tervalidasi masuk dalam *range* jarak yang telah ditentukan maka langkah selanjutnya adalah karyawan harus menginputkan data wajah kedalam sistem sebagai validasi kedua untuk melakukan presensi. Setelah melewati kedua tahap tersebut maka karyawan tercatat hadir secara sistem dan dapat menggunakan sistem presensi tersebut sebagai pelaporan kegiatan di hari itu.

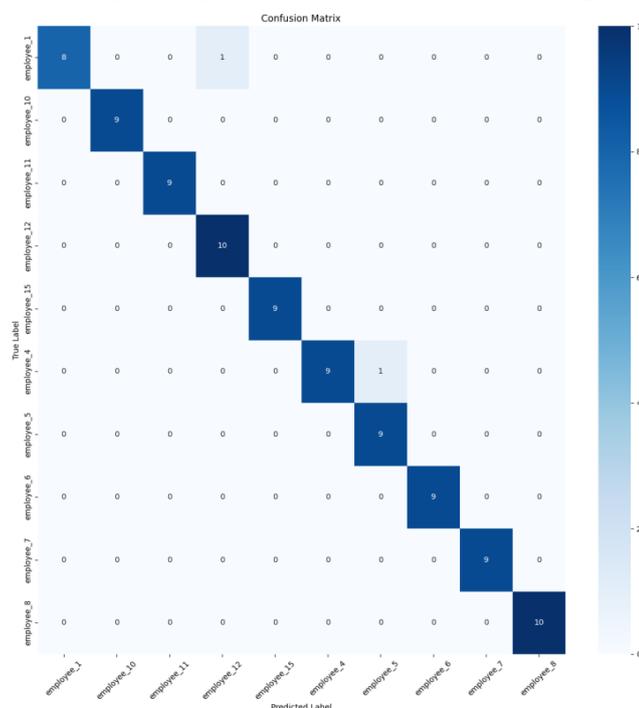
#### 3.2 Pengujian K-Nearest Neighbor (K-NN)

Tabel 2. Pengujian KNN

No.	Nama Karyawan	Data Wajah Setelah Proses Training
1.	Fahri	
2.	Andreas	

No.	Nama Karyawan	Data Wajah Setelah Proses Training
3.	Laurenza	

Pada Tabel 2 terdapat sampel data berupa gambar hasil proses *training* dari metode KNN. Sebanyak 90 data uji dievaluasi dan hasil prediksinya direpresentasikan dalam bentuk *confusion matrix*. *Confusion matrix* adalah sebuah tabel yang mengindikasikan jumlah data uji yang terklasifikasi dengan benar dan jumlah data uji yang terklasifikasi secara salah [7]. *Confusion matrix* menggambarkan seberapa baik model KNN dalam mengklasifikasikan data ke dalam kategori yang benar. *Confusion matrix* pada Gambar 4 merupakan hasil perhitungan dari 10 label yang masing-masing label memiliki nilai tersendiri. Terdapat 10 nilai *True Positive* dengan range nilai 8-10 dan 2 nilai *True Negative* dengan nilai 1.



Gambar 4. Confusion Matrix

Selanjutnya, Tabel 3 merupakan hasil pengukuran akurasi keseluruhan, presisi maupun *recall* masing-masing label atau kelas dengan model kostum KNN. Tabel ini menghasilkan akurasi *f1-score* dengan nilai 98% dan akurasi *support* sebesar 93%, *macro average* untuk presisi, *recall*, *f1-score* dengan nilai 98%, *support* sebesar 93% serta *weighted average* untuk presisi, *recall*, *f1-score* dengan nilai 98%, *support* sebesar 93%.

Tabel 3. Hasil Pengukuran

Label	Precision	Recall	F1-Score	Support
Rico (employee_5)	1.00	1.00	1.00	9
Daniel (employee_1)	1.00	0.89	0.94	9
Andreas (employee_11)	1.00	1.00	1.00	9
Fahri (employee_12)	0.91	1.00	0.95	10
Viko (employee_6)	1.00	1.00	1.00	9
Muthiara (employee_8)	1.00	1.00	1.00	10

Label	Precision	Recall	F1-Score	Support
Niken (employee_4)	1.00	0.90	0.95	10
Riki (employee_15)	1.00	1.00	1.00	9
Laurenza (employee_7)	1.00	1.00	1.00	9
Novinda (employee_10)	1.00	1.00	1.00	9
accuracy			0.98	93
macro avg	0.98	0.98	0.98	93
weighted avg	0.98	0.98	0.98	93

Berdasarkan Tabel 3 dan Gambar 4 hasil *confusion matrix* pengujian *custom* model KNN, maka perolehan nilai akurasi, presisi dan *recall* menggunakan rumus perhitungan sebagai berikut:

### 3.2.1 Akurasi Model

Nilai Akurasi Keseluruhan didapat dengan total akurasi dibagi banyaknya k [8]. TP dan TN berdasarkan Gambar 5 merupakan diagonal utama dari *confusion matrix*. Formula akurasi terdapat pada persamaan 1.

$$akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Akurasi = \frac{926}{928} = 0.99784$$

### 3.2.2 Presisi Model

Presisi merupakan perbandingan antara jumlah prediksi positif yang tepat dengan total hasil dari seluruh prediksi positif [9]. Formula presisi terdapat pada persamaan 2.

$$Presisi\ Keseluruhan = \frac{Total\ Presisi\ Seluruh\ Label}{Total\ Label} \quad (2)$$

$$Presisi = \frac{9.91}{10} = 0.991$$

### 3.2.3 Recall Model

Menurut Pendit, dalam pengertian *Recall* yang dikemukakan oleh Lancaster, merupakan rasio dari jumlah dokumen yang dapat ditemukan kembali melalui suatu proses pencarian informasi.[10]. Formula recall terdapat pada persamaan 3.

$$Recall\ Keseluruhan = \frac{Total\ Recall\ Seluruh\ Label}{Total\ Label} \quad (3)$$

$$Recall = \frac{9.79}{10} = 0.979$$

## 3.3 Pengujian Haversine Formula

Pengujian ini dilakukan dengan lokasi Wisma Bermis RSPAD Gatot Subroto dengan koordinat - 6,177430936922497, 106,83566061526238 dan radius maksimal 20 meter. Hasil dari pengujian *blackbox* tahap ini dapat dilihat pada Tabel 4.

Tabel 4. Black Box Haversine Formula

No	Skenario	Titik Lokasi	Hasil Diharapkan	Kesimpulan Akhir
1	presensi di lokasi yang ditentukan	lat - 6,1777268, lon 106,8355965	sistem mengizinkan untuk melakukan pengujian tahap kedua	sesuai <i>valid</i>
2	presensi di luar lokasi yang ditentukan	lat - 6,1780842, lon 106,8343466	sistem tidak mengizinkan untuk melakukan pengujian tahap kedua dan memberikan pesan "You are too far from the attendance area"	sesuai <i>valid</i>

No	Skenario	Titik Lokasi	Hasil Diharapkan	Kesimpulan Akhir
3	melakukan presensi masuk	lat - 6,1777268, lon 106,8355965	sistem menerima dan menampilkan pesan "Berhasil mengisi presensi kehadiran!"	sesuai <i>valid</i>
4	melakukan presensi pulang	lat - 6,1777268, lon 106,8355965	sistem menerima dan menampilkan pesan "Berhasil mengisi presensi kepulangan!"	sesuai <i>valid</i>
5	melakukan presensi saat tidak ada jadwal	lat - 6,1777268, lon 106,8355965	sistem memberikan pesan "tidak ada presensi untuk hari ini"	sesuai <i>valid</i>

Berdasarkan hasil *blackbox testing* maka diperoleh nilai akurasi, presisi, *recall* sebagai berikut:

$$\text{akurasi} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$= \frac{5}{5}$$

$$\text{akurasi} = 1$$

Selanjutnya untuk nilai presisi sebagai berikut:

$$\text{Presisi Keseluruhan} = \frac{\text{Total Presisi Seluruh Label}}{\text{Total Label}}$$

$$= \frac{5}{5}$$

$$\text{Presisi} = 1$$

Selanjutnya untuk nilai *recall* sebagai berikut:

$$\text{Recall Keseluruhan} = \frac{\text{Total Recall Seluruh Label}}{\text{Total Label}}$$

$$= \frac{5}{5}$$

$$\text{Presisi} = 1$$

#### 4 Kesimpulan

Berdasarkan penulisan penelitian yang telah dilakukan dapat disimpulkan bahwa penggunaan metode *Haversine Formula* dengan bantuan *google maps API* terbukti mampu mencegah karyawan untuk melakukan presensi di tempat yang tidak seharusnya. Begitu juga dengan algoritma metode *K-Nearest Neighbor* (KNN) dapat diimplementasikan dan berjalan dengan baik serta efektif dalam melakukan identifikasi citra wajah. Pada penelitian ini menghasilkan akurasi *f1-score* sebesar 0.98 dan akurasi *support* sebesar 93. Selain itu, dihasilkan juga nilai *macro avg* pada *precision* sebesar 0.98, nilai *macro avg* pada *recall* sebesar 0.98, nilai *macro avg* pada *f1-score* sebesar 0.98 dan nilai *macro avg* pada *support* sebesar 93. Nilai *macro avg* berguna ketika ada ketidak seimbangan jumlah sampel antara kelas-kelas yang ada.

#### Ucapan Terima Kasih

Peneliti mengucapkan banyak terimakasih kepada PT. Sistem Integrasi Medika, Ibu Herlina Suganda selaku Direktur Operasional, Bapak Andreas Brahantiyo selaku Koordinator Lapangan, Muhammad Akbar Bagaskoro dan Ramdhan Dadan selaku programmer, Mutiara Panghurina dan Aini Cahyaning Putri selaku implementor serta senior dan teman-teman rekan kerja lainnya yang telah mengizinkan, memberikan dukungan dan kontribusi dalam menyediakan data dan informasi yang diperlukan untuk keperluan penelitian ini.



### Daftar Pustaka

- [1] I. Kurniawan, T. Irfan, P. Studi Elektronika Industri, J. Teknik Elektro, P. Negeri Jakarta, and J. G. Siwabessy, "Sistem Presensi Berbasis Face Recognition," 2019, Diakses pada 24 Juli 2023.
- [2] F. Pratama and P. Putra, "Pengembangan Sistem Presensi Untuk Work From Home (WFH) dan Work From Office (WFO) Selama Pandemi Covid-19," 2022, Diakses pada 24 Juli 2023.
- [3] M. M. Baharuddin, H. Azis, and T. Hasanuddin, "Analisis Performa Metode K-Nearest Neighbor Untuk Identifikasi Jenis Kaca," *ILKOM Jurnal Ilmiah*, vol. 11, no. 3, pp. 269–274, Dec. 2019, doi: 10.33096/ilkom.v11i3.489.269-274, Diakses pada 24 Juli 2023.
- [4] M. M. Baharuddin, H. Azis, and T. Hasanuddin, "Analisis Performa Metode K-Nearest Neighbor Untuk Identifikasi Jenis Kaca," *ILKOM Jurnal Ilmiah*, vol. 11, no. 3, pp. 269–274, Dec. 2019, Diakses pada 24 Juli 2023, doi: 10.33096/ilkom.v11i3.489.269-274.
- [5] M. Kamal Syarifudin, R. Titi, K. Sari, and M. K. Syarifudin, "Perbandingan Metode Euclidean Distance dan Haversine Distance pada Aplikasi Sistem PPDB dan algoritma K-Means Untuk Menentukan Kebijakan Peraturan Zonasi," vol. 15, no. 4, pp. 1979–276, 2022, Diakses pada 24 Juli 2023, doi: 10.30998/faktorexacta.v15i4.14795.
- [6] S. H. Bariah, M. Irsad, and S. Putra, "Penerapan Metode Waterfall Pada Perancangan Sistem Informasi Pengolahan Data Nilai Siswa", Diakses pada 24 Juli 2023.
- [7] A. Indriani, "Klasifikasi Data Forum dengan menggunakan Metode Naïve Bayes Classifier." [Online], Diakses pada 24 Juli 2023, Available: [www.bluefame.com](http://www.bluefame.com)
- [8] M. Rangga, A. Nasution, and M. Hayaty, "Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter," *Jurnal Informatika*, vol. 6, no. 2, pp. 212–218, 2019, [Online], Diakses pada 24 Juli 2023, Available: <http://ejournal.bsi.ac.id/ejurnal/index.php/ji>
- [9] K. L. Kohsasih, Z. Situmorang, and I. Artikel, "Analisis Perbandingan Algoritma C4.5 Dan Naïve Bayes Dalam Memprediksi Penyakit Cerebrovascular," *Jurnal Informatika*, vol. 9, no. 1, pp. 13–17, 2022, [Online], Diakses pada 24 Juli 2023, Available: <http://ejournal.bsi.ac.id/ejurnal/index.php/ji>
- [10] : Martin and L. Nilawati, "Recall dan Precision Pada Sistem Temu Kembali Informasi Online Public Access Catalogue (OPAC) di Perpustakaan," vol. 21, no. 1, pp. 77–84, 2019, Diakses pada 24 Juli 2023, doi: 10.31294/p.v20i2.