Journal of System & Technology
ISSN: 3110-0171
Vol. 1, No. 1, Juni 2025

KLASIFIKASI INDEKS STANDAR PENCEMARAN UDARA MENGGUNAKAN *LOCAL MEAN* DAN *DISTANCE WEIGHTING* K-*NEAREST NEIGHBOR*

Nabily Rafif[1], Khairul Umam Syaliman[2]*
[1]Teknik Informatika, Politeknik Caltex Riau, Pekanbaru, Riau
[2]Teknik Informatika, Universitas Riau, Pekanbaru, Riau
rafifnabily@gmail.com
*Corresponding author: khairul.umam@lecturer.unri.ac.id

Abstract: Kualitas udara yang buruk dapat berdampak serius pada kesehatan manusia dan lingkungan. Di Indonesia, informasi kualitas udara disampaikan melalui Indeks Standar Pencemar Udara (ISPU), namun penyebaran informasi ini masih terbatas karena hanya menggunakan alat monitor di titik tertentu dan proses perhitungannya dilakukan secara manual. Penelitian ini menawarkan solusi dengan mengembangkan sistem klasifikasi status ISPU berbasis web menggunakan metode Local Mean dan Distance Weighting pada algoritma K-Nearest Neighbor (LMDWkNN). Sistem ini dirancang untuk mempercepat proses klasifikasi ISPU secara otomatis dan real-time. Pengujian terhadap model menunjukkan kinerja akurasi sebesar 96%, precision 98%, recall 82%, dan f1-score 89%. Sementara itu, pengujian sistem menggunakan Black Box Testing menunjukkan semua fitur berjalan sesuai harapan. Sistem ini dapat menjadi alat bantu yang efektif dalam menyampaikan informasi kualitas udara kepada masyarakat secara cepat dan akurat

Keywords—ISPU, Klasifikasi, LMDWkNN, Udara, Website.

I. PENDAHULUAN

Udara merupakan gabungan alami gas yang mengelilingi planet Bumi. Terdiri dari 78% nitrogen, 20% oksigen, 0,93% argon, 0,30% karbon dioksida, serta jenis gas lainnya. Udara yang bersih mempunyai sejumlah keuntungan bagi keberlangsungan hidup manusia, di antaranya menyehatkan sistem pernapasan, mengurangi risiko penyakit jangka panjang, memperpanjang usia, meningkatkan daya tahan tubuh dan fokus, dan meningkatkan suasana hati [1], [2].

Kementerian Lingkungan Hidup dan Kehutanan (KLHK) menggunakan stasiun pemantau otomatis kontinu yang dimilikinya untuk menyampaikan informasi tentang kualitas udara kepada masyarakat. Indeks Standar Pencemar Udara (ISPU) dipakai oleh KLHK untuk menunjukkan bahwa kondisi udara bisa dikategorikan sebagai yang bersih. Dalam ISPU, KLHK telah menetapkan 7 (tujuh) parameter pengukuran, yakni PM10, PM2.5, NO_2, SO_2, CO, O_3, dan HC [3].

Saat ini, informasi tentang ISPU ini sangat penting bagi masyarakat untuk mengetahui status kualitas udara di kota mereka, terutama di daerah yang rawan kebakaran hutan. Ini akan memungkinkan pemerintah pusat dan pemerintah daerah untuk memberikan peringatan kepada masyarakat. Pemerintah setempat hanya menggunakan alat monitor kualitas udara yang tersebar di beberapa titik untuk menyampaikan informasi ISPU ini kepada masyarakat. Akibatnya, tidak semua masyarakat dapat mengetahui status kualitas udara secara real-time. Untuk menentukan status ISPU, setiap parameter harus dihitung secara manual, yang memakan waktu, tetapi hasilnya biasanya harus disampaikan kepada masyarakat secepat mungkin, paling sedikit dua kali dalam satu hari [3]. Salah satu solusi yang diusulkan adalah menciptakan sistem klasifikasi untuk penentuan status ISPU.

Penelitian sebelumnya [1], [4], [5], algoritma Naïve Bayes dan modifikasi k-Nearest Neighbor seperti LMkNN dan DWkNN telah digunakan untuk klasifikasi ISPU dengan hasil akurasi antara 88% hingga 98%. Berdasarkan hal tersebut, penelitian ini mengusulkan pembangunan sistem klasifikasi ISPU berbasis website dan aplikasi mobile menggunakan gabungan algoritma *Local Mean* dan *Distance Weighting k-Nearest Neighbor* (LMDWkNN), untuk mempermudah pengambilan keputusan dan memberikan informasi *real-time* kepada masyarakat.

II. TINJAUAN PUSTAKA

A. *Penelitian Terdahulu*

Penelitian yang dilakukan oleh [1] menggunakan algoritma Naïve Bayes untuk klasifikasi kualitas udara dan menghasilkan akurasi sebesar 88%. Penelitian lain oleh [4] mengembangkan metode LMkNN dan DWkNN yang dikombinasikan untuk mencapai akurasi hingga 97,99%. Sementara itu [5] menerapkan metode serupa di kota Makassar dan menemukan ketidaksesuaian data antar wilayah.

Penelitian terbaru oleh [2] menunjukkan efektivitas metode ensemble learning dalam prediksi kualitas udara. [6] serta [7] juga berhasil mengimplementasikan metode k-NN dan SVM dalam prediksi ISPU secara akurat. Referensi lain seperti [8] juga menunjukkan keberhasilan k-NN dalam klasifikasi komentar. Metode Local Mean (LMkNN) dan Distance Weighting (DWkNN) yang digunakan dalam penelitian ini telah terbukti mampu meningkatkan performa model klasifikasi seperti dibuktikan dalam penelitian oleh [9] dan [10].

B. *Local Mean dan Distance Weighting k-Nearest Neighbor*

Local Mean k-Nearest Neighbor (LMkNN) merupakan suatu adaptasi pendekatan yang telah dimodifikasi. Pendekatan ini merupakan bentuk klasifikasi non-parametrik yang sederhana namun efektif, dan

memiliki ketahanan yang terbukti. Telah diuji dan dibuktikan bahwa pendekatan ini mampu meningkatkan hasil klasifikasi serta mengurangi dampak dari bias yang mungkin timbul, terutama pada dataset dengan jumlah data yang terbatas [4], [10].

Seperti metode termodifikasi lainnya, *Distance Weighting K-Nearest Neighbor* (DWkNN) mendefinisikan kategori data yang baru berdasarkan nilai bobot yang dihasilkan dari jarak antar data. Hal ini dapat mengatasi kesalahan klasifikasi yang timbul akibat pengabaian kedekatan data, karena terbukti efektif dalam mengurangi bias dan ketidakseimbangan dalam distribusi kumpulan data [4], [9].

Proses untuk menggabungkan kedua metode di atas akan dijelaskan dalam tahap-tahap berikut [9]:

1) Mengidentifikasi nilai k sebagai total dari tetangga terdekat.
2) Dengan menggunakan metode Euclidean, jarak antara masing-masing data testing ke masing-masing data training dapat dihitung dengan persamaan 1:

$$D(x,y) = ||x-y||_2 = \sqrt{\sum_{j=1}^{N}|x-y|^2} \quad (1)$$

3) Urutkan data berdasarkan jaraknya dimulai dari yang paling kecil hingga yang paling besar, dengan jumlah nilai sebanyak k dari masing-masing kelas.
4) Mendapatkan nilai bobot berdasarkan jarak yang memisahkan data dari masing-masing kelasnya menggunakan persamaan 2:

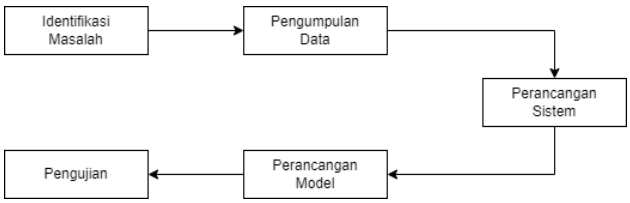
$$w_i = \frac{1}{d(x_q,x_i)} \quad (2)$$

5) Menghitung rata-rata bobot untuk masing-masing kelas data.
6) Menyeleksi kelas yang memiliki nilai tertinggi dari rata-rata bobot sebagai kelas yang akan digunakan untuk menguji data.

Dari tahapan di atas, untuk tahap 1 sampai tahap 3 merupakan kontribusi dari metode LMkNN, sedangkan tahap 4 sampai tahap 6 merupakan kontribusi dari metode DWkNN. Nilai k akan mengikuti aturan LMkNN yang mana nilai k adalah jumlah tetangga terdekat dari masing-masing kelas.

III. Metodologi Penelitian

Metodologi penelitian dalam pengembangan sistem klasifikasi ini dapat dilihat pada gambar 1.



Gambar 1. Metodologi Penelitian

Pada metodologi penelitian ini diawali dengan

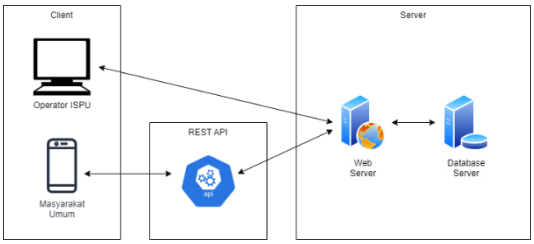
1) Identifikasi Masalah
   Masalah yang diidentifikasi adalah proses dalam penyampaian status ISPU masih menggunakan alat monitor kualitas udara yang hanya tersebar di beberapa titik tertentu sehingga tidak semua masyarakat dapat mengetahui status kualitas udara secara real-time dan juga proses penentuan status ISPU dilakukan secara manual, yang mana hal ini cukup memakan waktu.
2) Pengumpulan data
   Data ini diperoleh dari situs web resmi bank data portal https://data.jakarta.go.id/. Adapun rincian data dapat dilihat pada tabel 1.

Tabel 1. Pengumpulan Data

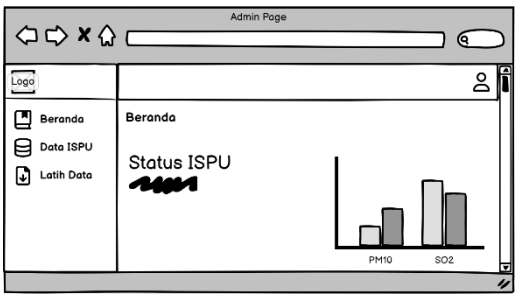
Tanggal	PM_{10}	SO_2	CO	...	Kategori
8/1/2020	69	22	16	...	TIDAK SEHAT
8/2/2020	91	27	26	...	TIDAK SEHAT
8/3/2020	95	26	19	...	SEDANG
...	...	...	...	...	...
8/31/2020	69	33	13	...	TIDAK SEHAT

Data yang diterapkan dalam studi ini adalah jenis data sekunder, yang berarti data ini diperoleh secara tidak langsung dari subjek penelitian. Kumpulan data ini terdiri dari 12 karakteristik, diantaranya tanggal, PM_{10}, PM_{25}, SO_2, CO, O_3, NO_2, HC, Max, Critical, dan Kategori.

3) Setelah model sudah selesai, tahap selanjutnya adalah perancangan sistem menggunakan framework Laravel dan Flutter [11], [12], [13], [14]. Adapun arsitektur sistem dapat dilihat pada gambar 2, dan perancangan dashboard dapat dilihat pada gambar 3.



Gambar 2. Arisitektur Sistem



Gambar 3. Rancangan Dashboard Sistem

4) Tahap terakhir dalam penelitian ini adalah pengujian terhadap sistem yang sudah dibangun [14]. Dalam proses pengujian ini yang pertama dilakukan pengujian akurasi terhadap modelnya, setalah itu barulah dilakukan pengujian terhadap sistem yang sudah dibangun.

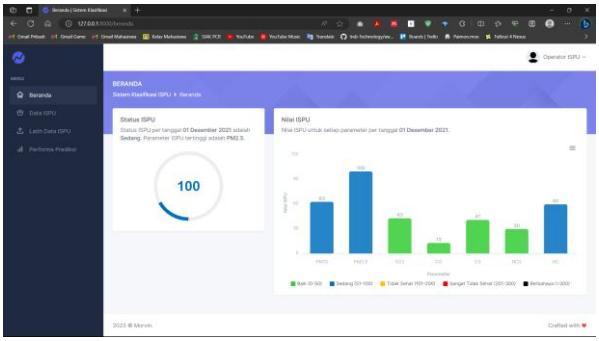
IV. Hasil dan Pembahasan

A. *Hasil Implementasi Sistem*

Selama implementasi sistem, pengembangan sistem klasifikasi ISPU ini menggunakan dua basis framework.

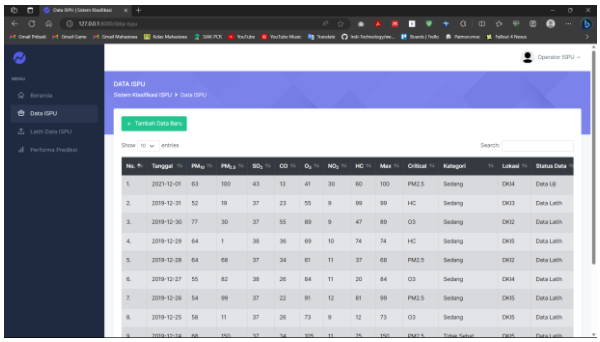
Journal of System & Technology
ISSN: 3110-0171
Vol. 1, No. 1, Juni 2025

Gambar 4 menampilkan halaman dashboard yang menampilkan informasi mengenai status Indeks Standar Pencemar Udara (ISPU) beserta nilai-nilai parameter yang terkait.



Gambar 4. Hasil Implementasi Dashboard

Gambar 5 menampilkan tampilan saat operator ISPU membuka opsi "Data ISPU" pada menu. Operator ISPU memiliki kemampuan untuk melihat rincian informasi mengenai data ISPU, menambahkan entri baru terkait ISPU, mengubah data ISPU yang sudah ada, serta menghapus data ISPU.



Gambar 5. Hasil Implementasi List Data ISPU

B. *Hasil Implementasi Model*

Berikut adalah hasil implementasi dari setiap proses yang diterapkan pada penelitian ini.

1) *Data Preprocessing*

Pada proses ini, data akan diambil dari database yaitu data ISPU yang telah ditambahkan dari tahun 2018 sampai dengan 2019 dengan total seluruh data yang ada berjumlah 710. Kemudian kolom data yang tidak digunakan untuk prediksi akan dibuang, adapun kolom data yang tidak digunakan yaitu kolom tanggal, max, critical, dan location. Setelah itu data akan dipisah menjadi fitur dan label. Adapun proses data preprocessing dapat dilihat pada gambar 6.

```php
// Mengambil data dari database yang stasus_data = Data Latih
$data = DataIspu::where('status_data', 'Data Latih')->get();

if ($data->isEmpty()) { // Jika data kosong
    return view('layout.performa.index', [
        'data' => '',
    ]);
} else { // Jika data tidak kosong
    // Memisah data menjadi fitur dan label
    $features = [];
    $labels = [];
    foreach ($data as $row) {
        $features[] = [$row->pm10, $row->pm25, $row->so2, $row->co, $row->o3, $row->no2, $row->hc];
        $labels[] = $row->categori;
    }

    // Menggabung kembali fitur dan label agar memudahkan saat membagi data
    $dataset = collect($features)->zip($labels);
```

Gambar 6. Data Preprocessing

2) *Modelling*

Pada proses ini, data akan dipecah menjadi dua bagian, yaitu data training dan data testing, dengan perbandingan 80:20. Langkah pembagian data dapat ditemukan dalam ilustrasi pada gambar 7.

```php
// Mengatur nilai seed untuk pengacakan
$seedValue = 42; // Nilai seed yang diinginkan
mt_srand($seedValue);

// Mengacak urutan elemen dalam koleksi
$shuffled = $dataset->shuffle();

// Menghitung jumlah sampel untuk set pelatihan dan set pengujian
$totalSamples = $shuffled->count();
$trainSize = (int) ($totalSamples * 0.8);

// Membagi koleksi menjadi set pelatihan dan set pengujian
$datasetParts = [
    'train' => $shuffled->take($trainSize),
    'test' => $shuffled->slice($trainSize)
];

// Memisahkan fitur-fitur dan label-label dari set pelatihan dan set pengujian
$trainData = $datasetParts['train']->pluck(0)->all();
$trainLabel = $datasetParts['train']->pluck(1)->all();
$testData = $datasetParts['test']->pluck(0)->all();
$testLabel = $datasetParts['test']->pluck(1)->all();
```

Gambar 7. Modelling

3) Evaluasi

Pada proses ini, dilakukan evaluasi hasil dari model yang telah dilatih dengan data yang sudah diproses pada tahap *Data Preprocessing*. Hasil evaluasi dari model menggunakan fungsi *Confusion Matrix* dan *Classification Report* dapat dilihat pada gambar 8 dan gambar 9.

```php
// Menghitung confusion matrix dan classification report
$lmdwknn = new LMDWKNN(7);
$lmdwknn->fit($trainData, $trainLabel);
$prediksi = $lmdwknn->predict($testData);
$confusion_matrix = $lmdwknn->confusion_matrix($testLabel, $prediksi);
$precision = $lmdwknn->precision($testLabel, $prediksi);
$recall = $lmdwknn->recall($testLabel, $prediksi);
$f1Score = $lmdwknn->f1Score($testLabel, $prediksi);
$average = $lmdwknn->average($testLabel, $prediksi);
$support = $lmdwknn->support($testLabel, $prediksi);
```

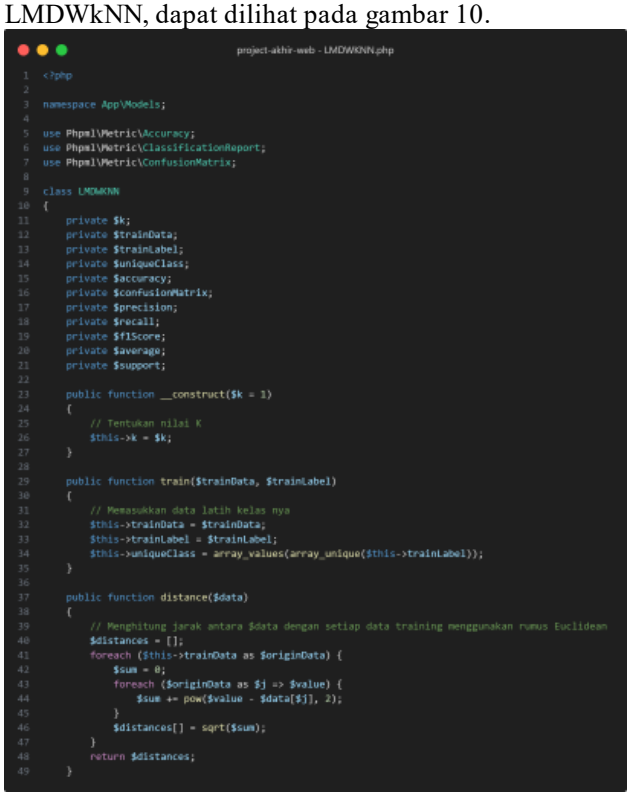
Gambar 8. Source Code Evaluasi Model

Gambar 9. Tampilan Hasil Evaluasi Model

C. *Deployment Model ke Website*

Selama perancangan sistem klasifikasi berlangsung, model LMDWkNN ini dibangun dalam bahasa pemrograman PHP yang mana bahasa pemrograman tersebut sudah terintegrasi dalam *framework Laravel* sehingga model LMDWkNN sudah dapat berjalan pada *back-end website*. Adapun proses *deployment* model ke *website* di antaranya yaitu pembuatan model LMDWkNN

dengan format *file* .php, perancangan *controller* untuk penggunaan model LMDWkNN, dan perancangan *route* berdasarkan fungsi yang telah dirancang pada *controller*. Berikut adalah potongan dari pembuatan model LMDWkNN, dapat dilihat pada gambar 10.



Gambar 10. Source Code Deployment Model

D. Pengujian Sistem dan Model

Untuk pengujian fungsional dari sistem dengan *Black Box Testing*, fungsi yang diujikan pada *website* yaitu fungsi *login*, mengelola data ISPU, dan melatih data baru, sedangkan fungsi yang diujikan pada aplikasi *mobile* yaitu fungsi melihat informasi ISPU. Adapun hasil dari pengujian *Black Box Testing* dapat dilihat pada tabel 2.

Tabel 2. Pengujian Blackbox

Fungsi yang diuji	Hasil yang diharapkan	Hasil Pengujian	Keterangan
...	...	...	...
Mengelola data ISPU	Data ISPU berhasil ditambah, diubah, dan dihapus	Berhasil	
...	...	...	...

Untuk pengujian model dengan perhitungan *Confusion Matrix*, digunakan 142 data dengan pembagian data yaitu: 5 data dengan label Baik, 67 data dengan label Sedang, 64 data dengan label Tidak Sehat, dan 6 data dengan label Sangat Tidak Sehat. Hasil dari perhitungan *Confusion Matrix* dapat ditemukan pada tabel 3.

Tabel 3. Confusion Matrix

	Predicted Value			
	Baik	**Sedang**	**Tidak Sehat**	**Sangat Tidak Sehat**
Baik	4	1	0	0
Sedang	0	67	0	0
Tidak Sehat	0	1	63	0
Sangat Tidak Sehat	0	0	3	3

Berdasarkan hasil pengujian *Confusion Matrix* (tabel 3) yang dilakukan, didapatkan kesimpulan akurasi dari model LMDWkNN ini sebesar 96%, yang artinya model ini dapat memprediksi status ISPU secara akurat. Hasil *precision* (atau presisi) dari model LMDWkNN ini sebesar 98%, yang artinya model ini secara presisi mampu menentukan prediksi kecocokan status ISPU. Hasil *recall* dari model LMDWkNN ini sebesar 82%, yang artinya model ini sudah baik dalam menggunakan data dan atribut untuk mengidentifikasi status ISPU. Hasil *f1-Score* dari model LMDWkNN ini sebesar 89%, yang artinya model ini memiliki keseimbangan yang baik antara *precision* dan *recall* dalam mengidentifikasi status ISPU.

V. KESIMPULAN

Berdasarkan hasil pengujian dan analisis yang telah dilakukan, maka dapat ditarik kesimpulan bahwa sistem klasifikasi status ISPU berbasis *website* dan aplikasi *mobile* telah berhasil dibangun. Hal ini dapat dilihat pada pengujian model menggunakan *Confusion Matrix* yang mana model ini bekerja dengan baik dalam menentukan status ISPU jika data ISPU baru ditambahkan dengan nilai kinerja rata-rata akurasi sebesar 96%, *precision* 98%, *recall* 82%, dan *f1-score* sebesar 89%. Sementara pada pengujian sistem menggunakan *Black Box Testing*, tidak ditemukan adanya kendala saat pengujian untuk setiap fungsionalitas pada *website*.

REFERENCES

[1] A. A. H. Kirono, I. Asror, and Y. F. A. Wibowo, "Klasifikasi Tingkat Kualitas Udara Dki Jakarta Menggunakan Algoritma Naïve Bayes," *e-Proceeding Eng.*, vol. 9, no. 3, p. 1962, 2022.

[2] R. Zhao *et al.*, "Human activities modulate greening patterns: a case study for southern Xinjiang in China based on long time series analysis," *Environ. Res. Lett.*, vol. 17, no. 4, p. 44012, Mar. 2022, doi: 10.1088/1748-9326/ac58a9.

[3] D. Chaniago, A. Zahara, and I. S. Ramadhani, "Indeks Standar Pencemar Udara (ISPU) Sebagai Informasi Mutu Udara Ambien Di Indonesia," ditppu.menlhk.go.id.

[4] Y. Yuliska and K. U. Syaliman, "Peningkatan Akurasi *k-Nearest Neighbor* Pada Data Index Standar Pencemaran Udara Kota Pekanbaru," *IT J. Res. Dev.*, vol. 5, no. 1, pp. 11–18, 2020, doi: 10.25299/itjrd.2020.vol5(1).4680.

[5] N. Aini, R. Ruktiari, R. Pratama, and A. Fitrah Buana, "Sistem Prediksi Tingkat Pencemaran Polusi Udara dengan Algoritma Naïve Bayes pada Kota Makassar," *Semin. Nas. Komun. dan Inform.*, pp. 83–90, 2019.

[6] A. C. Darmawan, "Pengembangan aplikasi berbasis web dengan python flask untuk klasifikasi data menggunakan metode decision tree C4. 5," 2023.

[7] R. R. A. Siregar, Z. U. Siregar, and R. Arianto, "Klasifikasi Sentiment Analysis Pada Komentar Peserta Diklat Menggunakan Metode *k-Nearest Neighbor*," *Kilat*, vol. 8, no. 1, pp. 81–92, 2019, doi: 10.33322/kilat.v8i1.421.

[8] Z. Pan, Y. Wang, and W. Ku, "A new k-harmonic nearest neighbor classifier based on the multi-local means," Expert Syst. Appl., vol. 67, pp. 115–125, 2017, doi: 10.1016/j.eswa.2016.09.031.

[9] K. U. Syaliman, E. B. Nababan, and O. S. Sitompul, "Improving the accuracy of *k-Nearest Neighbor* using *Local Mean* Based and Distance Weight," *J. Phys. Conf. Ser.*, vol. 978, no. 1, pp. 1–6, 2018, doi: 10.1088/1742-6596/978/1/012047.

[10] L. Afifah, "Algoritma *k-Nearest Neighbor* (kNN) untuk Klasifikasi," ilmudatapy.com.

[11] T. Rismawan, A. W. Irawan, W. Prabowo, and S. Kusumadewi, "Sistem Pendukung Keputusan Berbasis Pocket PC Sebagai Penentu Status Gizi Menggunakan Metode k-NN (*k-Nearest Neighbor*)," *Teknoin*, vol. 13, no. 2, pp. 18–23, 2008, doi: 10.20885/teknoin.vol13.iss2.art5.

[12] A. A. Wahid, "Analisis Metode Waterfall Untuk Pengembangan Sistem Informasi," *J. Ilmu-ilmu Inform. dan Manaj. STMIK*, no.

November, pp. 1–5, 2020.
[13] R. Juliarto, “Contoh Use Case Diagram Lengkap dengan Penjelasannya,” dicoding.com.
[14] T. S. Jaya, “Pengujian Aplikasi dengan Metode Blackbox Testing Boundary Value Analysis,” *J. Inform. Pengemb. IT*, vol. 3, no. 2, pp. 45–46, 2018.

Nabily Rafif, S.Tr.Kom, merupakan alumni dari Politeknik Caltex Riau dengan latar belakang pendidikan Teknik Informatika. Pernah terlibat dalam penelitian terkait komputasi yang berfokus pada machine learning.

Khairul Umam Syaliman adalah anak kedua dari empat bersaudara, lahir pada tanggal 21 Juni 1992 di Perawang, Kecamatan Tualang, Kabupaten Siak. Ia merupakan putra dari pasangan Lukman dan Syalinurwathy. Ia menyelesaikan pendidikan Magister (S2) pada Program Studi Teknik Informatika di Universitas Sumatera Utara (USU). Pada tahun 2019 hingga 2023, ia pernah menjadi dosen tetap pada Program Studi Teknik Informatika, Politeknik Caltex Riau. Sejak tahun 2024 telah aktif mengajar pada Program Studi Teknik Informatika, Universitas Riau (UNRI). Selain itu saat ini juga aktif meneliti dan menulis karya ilmiah pada bidang *Data Mining*, *Artificial Intelligence* (*Machine Learning*, *Deep Learning*), *Natural Languange Processing* (NLP), *Computer Vision*, dan *Software Engineering*.