

Implementasi Metode *Decision Tree* pada Sistem Prediksi Status Gizi Balita

Dasilva Nike Aria Kurniawan¹, Maryam^{2*}

^{1,2}Program Studi Informatika, Universitas Muhammadiyah Surakarta, Jawa Tengah, Indonesia
e-mail: l200190092@student.ums.ac.id¹, maryam@ums.ac.id²

Abstract

Examination of the nutritional status of toddlers is one way to monitor and identify toddlers who are at risk of experiencing nutritional problems. Implementation of the method in predicting nutritional status using relevant parameters. The Decision Tree method is used as a predictive model, using a dataset with parameters such as age, sex, height, weight, and nutritional status as labels. At the mining stage, data processing starts from preprocessing, namely the cleansing process to clean up incorrect data and data transformation to change the data type so that it is easy to process during the classification process. Furthermore, the Decision Tree model will be trained, tested and measured based on accuracy. The model is described in the form of a decision tree so that it can be used as a rule in system implementation. The implementation results provide accurate predictions with an accuracy value of 92.73%. The prediction system is designed to assist health workers in supporting decisions on predicting the nutritional status of children under five, as well as facilitating the community to carry out independent checks. This prediction can help identify toddlers at risk of nutritional disorders so that early intervention steps can be taken appropriately.

Keywords: Toddlers, Decision Tree, Prediction System, Nutritional Status

Abstrak

Pemeriksaan status gizi balita merupakan salah satu cara untuk memantau dan mengidentifikasi balita yang beresiko mengalami masalah gizi. Implementasi metode dalam melakukan prediksi status gizi menggunakan parameter yang relevan. Metode Decision Tree digunakan sebagai model prediktif, menggunakan dataset dengan parameter berupa umur, jenis kelamin, tinggi badan, berat badan, serta status gizi sebagai label. Pada tahap mining, pengolahan data dimulai dari preprocessing yaitu proses cleansing untuk membersihkan data yang salah dan transformasi data untuk merubah tipe data agar mudah diolah saat proses klasifikasi. Selanjutnya, model Decision Tree akan dilatih, diuji dan diukur berdasarkan akurasi. Model digambarkan dalam bentuk pohon keputusan sehingga dapat digunakan sebagai rule dalam implementasi sistem. Hasil implementasi memberikan prediksi yang akurat dengan nilai akurasi sebesar 92,73%. Sistem prediksi dirancang untuk membantu tenaga kesehatan dalam mendukung keputusan hasil prediksi status gizi balita, sekaligus memfasilitasi masyarakat untuk melakukan pemeriksaan secara mandiri. Prediksi ini dapat membantu mengidentifikasi balita dengan risiko kelainan gizi sehingga langkah-langkah intervensi dini dapat diambil secara tepat.

Kata kunci: Balita, Decision Tree, Sistem Prediksi Status Gizi

1. PENDAHULUAN

Anak balita (1-5 tahun) merupakan kelompok umur yang paling rentan kekurangan gizi [1]. Masalah gizi masih didominasi oleh keadaan kurang gizi seperti anemia besi, gangguan akibat kurang yodium, kurang vitamin A dan kekurangan energi protein (KEP) [2]. Dampak yang terjadi apabila gizi balita tidak terpenuhi akan berpengaruh terhadap tumbuh kembang balita, menghambat perkembangan kognitif, dan meningkatkan resiko kematian balita. Balita yang kekurangan gizi tidak mampu membentuk antibodi terhadap penyakit infeksi



sebagai akibatnya anak-anak sering kali terkena penyakit sehingga mengganggu pertumbuhannya. Penyebab Balita memiliki gizi buruk ialah kemiskinan sehingga akses pangan si balita terganggu dan juga ketidaktahuannya orang tua karena kurang Pendidikan sehingga pengetahuan gizi rendah. Status Gizi adalah merupakan suatu kondisi atau keadaan tubuh yang diakibatkan oleh makanan yang dikonsumsi, penyerapan dan penggunaan makanan oleh tubuh [3]. Oleh karena itu, pemeriksaan status gizi diperlukan untuk memantau status gizi balita sehingga dapat melakukan penanganan yang tepat. Hal ini dapat dibantu dengan adanya sistem yang dapat membantu proses prediksi status gizi balita apakah memiliki gizi yang seimbang atau tidak.

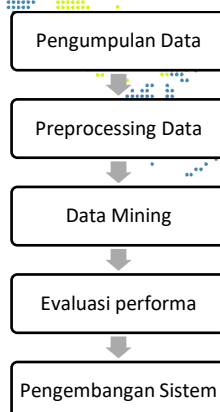
Penelitian pada prediksi status Gizi Balita dilakukan dengan berbagai macam metode. Metode *Decision Tree* digunakan dan memiliki tingkat akurasi 100% [4]. Metode Naïve Bayes dengan nilai akurasi 80.60% [5]. Metode *K-Nearest Neighbor* memperoleh akurasi sebesar 68% [6]. Berdasarkan hasil Riset Kesehatan Dasar, bahwa status gizi balita memuat, Umur Balita, Jenis Kelamin, berat badan menurut umur (BB/U), dan tinggi badan menurut umur (TB/U) merupakan indikator yang umum digunakan karena memiliki keunggulan dan lebih mudah dimengerti oleh masyarakat umum [7].

Metode *Decision Tree* ini populer karena mampu melakukan klasifikasi sekaligus menunjukkan hubungan antar atribut. Berbagai macam algoritma yang dapat membangun suatu *decision tree* salah satunya C4.5. Algoritma C4.5 merupakan salah satu algoritma yang digunakan untuk melakukan klasifikasi atau segmentasi atau pengelompokan dan bersifat prediktif. Pada pelayanan kesehatan masyarakat semisal Puskesmas, proses pemeriksaan status gizi balita rutin dilakukan, dengan hasil beragam. Hal ini terjadi karena balita merupakan kelompok yang rentan menderita kelainan gizi. Proses pemeriksaan dapat dipermudah dengan adanya suatu sistem untuk melakukan pemeriksaan secara mandiri oleh masyarakat, cukup dengan menginputkan data anak. Sistem ini sekaligus membantu tenaga kesehatan dalam memutuskan hasil pemeriksaan status gizi balita.

Berdasarkan permasalahan tersebut, model klasifikasi yang tepat diperlukan untuk membangun sistem prediksi status gizi balita. Metode yang digunakan adalah *Decision Tree* C4.5 dengan melibatkan data primer yang diperoleh dari puskesmas dengan 4 parameter berupa usia, jenis kelamin, tinggi badan, dan berat badan. Proses pengembangan sistem diimplementasikan berbasis website sehingga dapat membantu tenaga kesehatan dalam proses pengambilan keputusan serta dapat membantu masyarakat umum dalam melakukan pemeriksaan secara mandiri, dan kemudian memperoleh info penanganan yang tepat sesuai hasil prediksi.

2. METODOLOGI PENELITIAN

Penelitian ini memiliki beberapa tahap yang diperlukan untuk menghasilkan sistem prediksi status gizi. Alur pada metode penelitian terdapat pada Gambar 1.



Gambar 1. Tahapan penelitian

2.1. Pengumpulan Data

Pengumpulan data diperoleh menggunakan metode wawancara dan obeservasi dengan petugas pada puskesmas Sibela, Kelurahan Mojosongo, Kecamatan Jebres, Jawa Tengah. Hasil perolehan data untuk penelitian berupa data primer dari pemeriksaan balita yang dilakukan pada tahun 2021.

2.2. Preprocessing Data

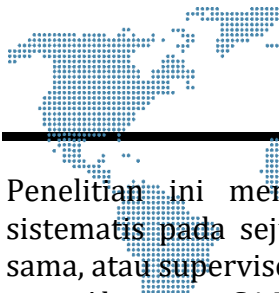
Data yang diperoleh kemudian dilakukan integrasi berupa pengelompokkan data dengan menggunakan 4 parameter yaitu: Umur, Jenis Kelamin, Berat badan menurut umur (BB/U), dan Tinggi badan menurut umur (TB/U). Hasil status gizi dinyatakan dengan 3 label yaitu status gizi kurang, gizi baik, dan gizi lebih. Proses *cleansing* dilakukan untuk membersihkan data yang salah, tidak lengkap, atau data yang hilang. Kemudian dilakukan transformasi data untuk merubah tipe data agar mudah diolah saat proses klasifikasi. Tabel 1 menunjukkan data pemeriksaan balita sebanyak 55 *record*.

Tabel 1. Data Pemeriksaan Balita

No	Jenis Kelamin	Umur	BB/U	TB/U	Status Gizi
1	Laki-Laki	1 th	6,5	68,6	Gizi Kurang
2	Laki-Laki	1 th	6,7	70,5	Gizi Kurang
3	Laki-Laki	1 th	6,5	68,6	Gizi Kurang
4	Laki-Laki	1 th	8,6	73,4	Gizi Baik
5	Perempuan	1 th	8,9	71,4	Gizi Baik
..
54	Perempuan	5 th	29,5	95,2	Gizi Lebih
55	Perempuan	5 th	29,5	123,7	Gizi Lebih

2.3. Data Mining

Data Mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual [9]. Data mining sering disebut juga *Knowledge Discovery in Database* (KDD) dan memiliki beberapa teknik, antara lain adalah *classification* [10].



Penelitian ini menggunakan teknik klasifikasi yaitu pengelompokan yang sistematis pada sejumlah objek ke dalam kelas berdasarkan karakteristik yang sama, atau supervised learning dengan ciri berupa data yang memiliki label.

Algoritma C4.5 yaitu algoritma yang digunakan untuk membangun *decision tree* (pohon keputusan) [11]. Metode ini mampu mengambil keputusan dengan cara sederhana, sehingga mudah dipahami. Kelebihan dari metode ini adalah hasilnya dapat digambarkan bentuk pohon keputusan sehingga dapat diamati langsung [8]. Pohon keputusan dapat membagi kumpulan data besar menjadi himpunan-himpunan record yang lebih kecil dengan menerapkan serangkaian aturan keputusan. Rumus Algoritma C4.5 terbagi menjadi 2 formula, formula (1) digunakan untuk mencari nilai gain dan formula (2) mencari nilai entropy.

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (1)$$

Keterangan :

S = Himpunan Kasus

A = Fitur

n = Jumlah partisi atribut A

|S_i| = Proporsi S_i terhadap S

|S| = Jumlah kasus dalam S

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i \quad (2)$$

Keterangan :

S = Himpunan Kasus

A = Fitur

N = Jumlah Partisi S

Pi = Proporsi dari S_i terhadap S

2.4. Evaluasi Performa

Evaluasi performa menggunakan *Confusion Matrix* untuk menguji seberapa akurat metode klasifikasi C4.5. *Confusion Matrix*, dikenal sebagai table kontigensi yang memungkinkan visualisasi kinerja algoritma pembelajaran. Setiap Kolom matriks mewakili contoh dalam kelas yang di prediksi, sementara setiap baris mewakili contoh dalam kelas yang sebenarnya [12]. *Confusion Matrix* dapat dilihat pada Tabel 2.

Tabel 2. *Confusion Matrix*

Kelas Prediksi	Kelas Aktual		
		Positif	Negatif
	Positif	True Positif (TP)	False Positif (FP)
	Negatif	False Negatif (FN)	True Negatif (TN)

Nilai Akurasi dapat dihitung dengan menggunakan formula (3) yang didefinisikan sebagai berikut :

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

Keterangan :

TP : Jumlah kelas yang diprediksi positif dan kelas aktualnya adalah positif

TN : Jumlah kelas yang diprediksi negatif dan kelas aktualnya adalah negatif

FP : Jumlah kelas yang diprediksi positif tetapi kelas aktualnya adalah negatif

FN : Jumlah kelas yang diprediksi negatif tetapi kelas aktualnya adalah positif

2.5. Pengembangan Sistem

Pengembangan dilakukan dengan sistem SDLC (*Software Development Life Cycle*) dan menggunakan metode waterfall. Berisi rangkaian aktivitas proses yang teratur dan disajikan dalam proses yang terpisah, seperti spesifikasi kebutuhan, implementasi desain perangkat lunak, uji coba, dan *maintenance* [13]-[14]. Model air terjun (*Waterfall*) ini menyediakan pendekatan alur hidup secara sekuensial [15].

2.5.1. Analisis Kebutuhan

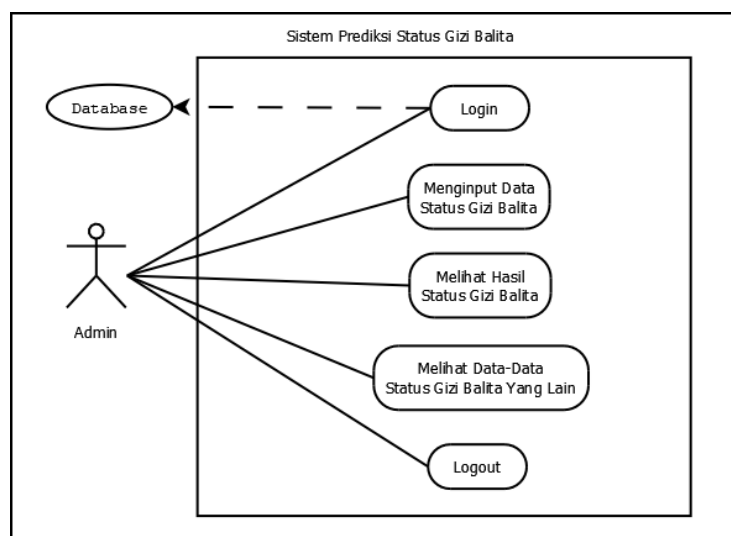
Sistem memiliki beberapa fitur antaranya melakukan login dan logout, menginput data balita, menginput data kriteria, melihat hasil data status gizi balita, dan melihat hasil prediksi semua data balita.

2.5.2. Design

Pembuatan desain meliputi *Use Case Diagram*, dan *Activity Diagram*.

a. Use Case Diagram

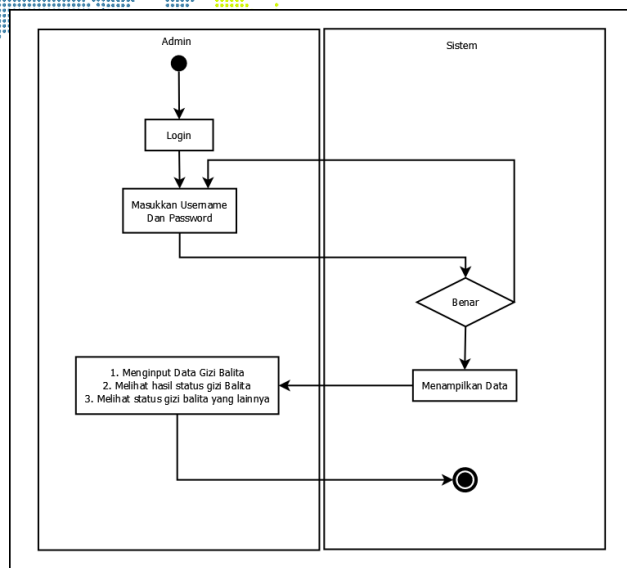
Use Case Diagram mendefinisikan fungsionalitas dan grafis dari sebuah sistem dari segi actor, *use case*, dan relasi [16]. Pada Gambar 2 menunjukkan bahwa admin dapat melakukan login, menginput data gizi balita, dan melihat hasil prediksi, serta data semua balita.



Gambar 2. *Use Case Diagram*

b. Diagram Activity

Activity Diagram adalah diagram untuk memodelkan aspek dinamis dari suatu sistem [17]. Pada Gambar 3 *Diagram Activity Admin* diawali dengan login, jika berhasil maka akan diarahkan ke halaman admin.



Gambar 3. Diagram Activity Admin

2.5.3. Implementasi

Pengembangan sistem menggunakan Bahasa pemrograman *Python*, *Python* menjadi Bahasa pemrograman yang dipakai secara luas dalam industry dan Pendidikan karena sederhana, ringkas, sintak intuitif dan memiliki Pustaka yang luas. Rule yang digunakan dalam membuat sistem merupakan hasil dari prediksi menggunakan algoritma C4.5.

2.5.4. Testing

Pengujian sistem akan menggunakan *Black-Box*, *Black-Box* adalah tes yang berfokus pada spesifikasi fungsional dari perangkat lunak, penguji dapat menentukan serangkaian kondisi input dan menjalankan tes untuk setiap fungsi pada program [19].

3. HASIL DAN PEMBAHASAN

3.1. Hasil Pemodelan C4.5

Kriteria yang digunakan antara lain : Berat Badan menurut Umur (BB/U), Tinggi Badan menurut Umur (TB/U), dan Status Gizi sebagai label. Setelah data diproses menggunakan algoritma C4.5 maka menghasilkan rule yang akan di implementasikan ke program. Rule yang dihasilkan sebagai berikut:

```
Status Gizi Baik
  If Jenis Kelamin = Laki-Laki
    If Berat Badan = Ringan
      Status Gizi = Baik
    Else If Jenis Kelamin = Perempuan
      If Berat Badan = Ringan
        Status Gizi = Baik
      Status Gizi = Kurang
    Else If Berat Badan = Ringan
      Status Gizi = Kurang
  Status Gizi = Lebih
  Else If Berat Badan = Berat
    Status Gizi = Lebih
```

3.2. Hasil Prediksi

Hasil prediksi merupakan proses untuk menemukan pola dari data dengan menggunakan beberapa variable [21]. Berdasar data pemeriksaan balita, terdapat status gizi kurang dengan total 3 balita, gizi baik 42 balita, dan gizi lebih dengan 6 balita. Setelah implementasi metode C4.5 maka diperoleh performance sesuai Tabel 3.

Tabel 3. Hasil *Performance*

Performance Vector				
Accuracy: 92.73%				
	Gizi Kurang	Gizi Baik	Gizi Lebih	class precision
Gizi Kurang	3	4	0	42.86%
Gizi Baik	0	42	0	100.00%
Gizi Lebih	0	0	6	100.00%
class recall	100.00%	91.30%	100.00%	

Berdasar table *performance* diperoleh akurasi sebesar 92.73% dengan rincian sebagai berikut :

- Status gizi kurang terdefinisi benar sebagai gizi buruk sebanyak 3 data.
- Status gizi kurang terdefinisi salah sebagai gizi baik sebanyak 4 data.
- Status gizi baik terdefinisi benar sebagai gizi baik sebanyak 42 data.
- Status gizi lebih terdefinisi benar sebagai gizi lebih sebanyak 6 data.

3.3. Implementasi Sistem

Penelitian ini menghasilkan Sistem Prediksi Status Gizi Balita dengan Metode Decision Tree C4.5 yang mampu memberikan informasi mengenai status gizi balita. Gambar 4 menunjukkan halaman prediksi status gizi balita yang menampilkan status gizi balita sesuai data anak yang dimasukkan. Langkah awal yaitu user mengisi data pada menu data identitas, data kriteria, dan data anak. Pengelolaan data pada sisi admin dapat melakukan proses tambah, update, dan hapus data. Pada menu Data Identitas Anak, tersedia masukan berupa NIK, Nama, Tanggal Lahir, Jenis Kelamin dan Alamat Rumah. Sedangkan pada menu Data anak, berupa masukan pengisian data gizi berupa jenis kelamin, umur, berat badan, dan tinggi badan anak. Proses dilanjutkan untuk melakukan prediksi dan memproses data serta menghasilkan prediksi berupa gizi buruk, gizi baik, atau gizi lebih.

No	NIK	Nama	Jenis Kelamin	Umur	Hasil (BB/U)	Hasil (TB/U)	Gizi
1	71021	Surya	Laki-Laki	2	12	85	Baik
2	170821	Putri	Perempuan	2	8	75	Buruk
3	310522	Joko	Laki-Laki	1	13	75	Baik

Gambar 4. Halaman Prediksi Status Gizi

3.4. Pengujian Metode *Black-Box*

Sistem Prediksi kemudian dilakukan pengujian *Black-Box* dan menunjukkan hasil valid yang ditunjukkan pada Tabel 4.

Tabel 4. Hasil Uji *Black-Box*

Aksi	Skenario Pengujian	Hasil yang diharapkan	Status
Login	input username dan Password benar.	Menampilkan halaman Dashboard	Valid
	input username dan Password salah	Menampilkan Halaman login	Valid
Data Identitas	Mengelola identitas data balita	Input, edit, dan hapus data berhasil	Valid
Data Kriteria	Mengelola data kriteria	Input, edit, dan hapus data berhasil	Valid
Data Anak	Menginputkan data sesuai kriteria prediksi	Input data berhasil	Valid
Proses Prediksi	Melakukan proses Prediksi	Menampilkan halaman hasil Prediksi Status Gizi pada data anak	Valid

4. SIMPULAN

Berdasarkan hasil penelitian algoritma C4.5 dapat diterapkan dan handal dalam menentukan status gizi balita. Penggunaan 4 parameter antara lain: Jenis Kelamin, Umur, Berat badan menurut umur (BB/U), dan Tinggi badan menurut umur (TB/U) menjadi atribut yang paling berpengaruh terhadap status gizi balita. Hasil klasifikasi memiliki *performance* yang unggul sebesar 92,73%. Selanjutnya, rule yang diperoleh diterapkan kedalam sistem. Uji sistem menggunakan blackbox dinyatakan valid, sehingga sistem dapat digunakan sebagai alat bantu dalam proses prediksi status gizi balita.

DAFTAR PUSTAKA

- [1] N. R. Fauzia, N. M. A. Sukmandari, and K. Y. Triana, "Hubungan Status Pekerjaan Ibu Dengan Status Gizi Balita," *J. Cent. Res. Publ. Midwifery Nurs.*, vol. 3, no. 1, pp. 28–32, 2019, doi: 10.36474/caring.v3i1.101.
- [2] L. Suryani, S. Payung, and N. Pekanbaru, "Faktor Yang Mempengaruhi Status Gizi Balita Di Wilayah Kerja Puskesmas Payung Sekaki," *JOMIS (Journal Midwifery Sci.*, vol. 1, no. 2, pp. 47–53, 2017,
- [3] P. Studi *et al.*, "Status Gizi Balita Di Dinas Kesehatan".
- [4] A. Prasetio, "Simulasi Penerapan Metode Decision Tree (C4.5) Pada Penentuan Status Gizi Balita," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 4, no. 3, pp. 209–214, 2021, doi: 10.32672/jnkti.v4i3.2983.
- [5] R. Setiawan and A. Triayudi, "Klasifikasi Status Gizi Balita Menggunakan Naïve Bayes dan K-Nearest Neighbor Berbasis Web," *J. Media Inform. Budidarma*, vol. 6, no. 2, p. 777, 2022, doi: 10.30865/mib.v6i2.3566.
- [6] E. T. Lestari and J. Adhiva, "Implementation Naive Bayes Classifier Algorithm and K-Nearest Neighbor For Obesity Nutritional Status of Children with Disabilities Implementasi Algoritma Naive Bayes Classifier dan K-Nearest Neighbor Untuk Klasifikasi Status Gizi Obesitas Anak Disabilitas," *SENTIMAS Semin. Nasional Penelit. dan Penabdian Masy.*, pp. 1–11, 2022.



- [7] K. Riskerdas, "Hasil Utama Riset Kesehatan Dasar (RISKEDAS)," *J. Phys. A Math. Theor.*, vol. 8, no. 44, pp. 1–200, 2018.
- [8] F. Gorunescu, *Data Mining: Concepts, models, and techniques (Softcover)*. 2011.
- [7] D. Setianingsih and R. F. Hakim, "Penerapan Data Mining dalam Analisis Kejadian Tanah Longsor di Indonesia dengan Menggunakan Association Rule Algoritma Apriori," *Pros. Semin. Nas. Mat. dan Pendidik. Mat. UMS 2015*, pp. 731–741, 2015.
- [8] Lindawati, "Data Mining Dengan Teknik Clustering Dalam Pengklasifikasian Data Mahasiswa Studi Kasus Prediksi Lama Studi Mahasiswa Universitas Bina Nusantara," *Univ. Stuttgart*, vol. 2008, no. semnasIF, pp. 174–180, 2008.
- [9] R. Kurniasari and A. Fatmawati, "Penerapan Algoritma C4.5 Untuk Penjurusan Siswa Sekolah Menengah Atas," *Komputa J. Ilm. Komput. dan Inform.*, vol. 8, no. 1, pp. 19–27, 2019, doi: 10.34010/komputa.v8i1.3045.
- [10] V. M. Patro and M. Ranjan Patra, "Augmenting Weighted Average with Confusion Matrix to Enhance Classification Accuracy," *Trans. Mach. Learn. Artif. Intell.*, vol. 2, no. 4, 2014, doi: 10.14738/tmlai.24.328.
- [11] E. Sudarmilah and P. A. Wibowo, "Aplikasi Augmented Reality Game Edukasi Untuk Pengenalan Organ Tubuh Manusia," *Khazanah Inform. J. Ilmu Komput. dan Inform.*, vol. 2, no. 1, pp. 20–25, 2016, doi: 10.23917/khif.v2i1.1945.
- [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] M. Susilo, "Rancang Bangun Website Toko Online Menggunakan Metode Waterfall," *InfoTekJar (Jurnal Nas. Inform. dan Teknol. Jaringan)*, vol. 2, no. 2, pp. 98–105, 2018, doi: 10.30743/infotekjar.v2i2.171.
- [14] M. N. Arifin and D. Siahaan, "Structural and Semantic Similarity Measurement of UML Use Case Diagram," *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 11, no. 2, p. 88, 2020, doi: 10.24843/lkjiti.2020.v11.i02.p03.
- [15] B. Khafid and D. A. P. Putri, "Pesma Apps as Android-based Integrated Applications for Mahasantri Pesma KH Mas Mansur UMS," *Khazanah Inform. J. Ilmu Komput. dan Inform.*, vol. 6, no. 2, pp. 95–102, 2020, doi: 10.23917/khif.v6i2.10494.
- [16] R. B. Susanto, P. S. Informatika, F. Komunikasi, D. A. N. Informatika, and U. M. Surakarta, "Gudang (Studi Kasus Di Yulian Aquatic)," 2022.
- [17] T. Hidayat and M. Muttaqin, "Pengujian sistem informasi pendaftaran dan pembayaran wisuda online menggunakan black box testing dengan metode equivalence partitioning dan boundary value analysis," *J. Tek. Inform. UNIS*, vol. 6, no. 1, pp. 2252–5351, 2018, doi: 10.33005/jifosi.v2i2.326.
- [18] M. Maryam and H. W. Ariono, "Sistem Pakar Pengklasifikasi Stadium Kanker Serviks Berbasis Mobile Menggunakan Metode Decision Tree," *J. Kaji. Ilm.*, vol. 22, no. 3, pp. 267–278, 2022, doi: 10.31599/jki.v22i3.1368.
- [19] M. V. Ruma, Yohanes, "Klasifikasi Status Gizi Balita Menggunakan Metode Naive Bayes Classification," p. 133, 2020.