

SISTEM PAKAR DIAGNOSIS BALITA: FORWARD CHAINING DAN FUZZY MATCHING UNTUK VALIDASI GEJALA

Iwan Setiawan

D4 Logistik Niaga-EL, Sekolah Vokasi Universitas Logistik dan Bisnis Internasional

email: iwan.setiawan@ulbi.ac.id

Abstrak

Mendiagnosis penyakit pada balita sering kali menjadi tantangan bagi tenaga kesehatan, terutama ketika gejala yang muncul, seperti demam, batuk, dan pilek, memiliki kemiripan. Kesulitan ini semakin kompleks dengan adanya variasi cara orang tua mendeskripsikan gejala, seperti menyebut "pilek" dengan istilah "flu ringan" atau "hidung tersumbat." Perbedaan semantik ini sering kali menyebabkan ketidakpastian yang memengaruhi akurasi diagnosis, terutama jika hanya mengandalkan metode konvensional. Penelitian ini menawarkan solusi berupa pengembangan sistem pakar yang memadukan metode **Forward Chaining** dan **Fuzzy Matching** untuk meningkatkan keakuratan diagnosis. Teknologi **Natural Language Processing (NLP)** digunakan untuk mengenali gejala dari deskripsi teks yang diberikan pengguna, sementara **Fuzzy Matching** dirancang agar sistem mampu memproses input dengan toleransi terhadap kesalahan kecil dalam penulisan. Tingkat kepercayaan diagnosis dihitung menggunakan **Certainty Factor (CF)**, yang memberikan bobot berdasarkan kesesuaian gejala. Pengujian menunjukkan bahwa sistem ini memiliki akurasi hingga 91% dan waktu pemrosesan yang singkat, rata-rata kurang dari 1 detik. Sistem ini diharapkan dapat menjadi alat bantu yang bermanfaat bagi tenaga kesehatan, terutama di daerah dengan akses terbatas ke dokter spesialis, sehingga mempercepat pengambilan keputusan yang tepat untuk kesehatan balita

Kata Kunci: forward chaining, fuzzy matching, diagnosis penyakit, balita, NLP

Abstract

Diagnosing infections in toddlers typically provides a problem for healthcare workers, especially when the symptoms that appear, such as fever, cough, and runny nose, are similar. This challenge becomes even more complex with the variance in how parents describe symptoms, such as referring to a "cold" as "mild flu" or "stuffy nose." These lexical distinctions often lead to doubts that impair the accuracy of diagnosis, especially if relying entirely on conventional procedures. This research proposes a solution in the form of constructing an expert system that combines Forward Chaining and Fuzzy Matching approaches to improve diagnostic accuracy. Natural Language Processing (NLP) technology is utilized to distinguish symptoms from the text description provided by the user, while Fuzzy Matching is built so that the system can analyze input with tolerance for slight writing errors. The confidence level of the diagnosis is derived using the Certainty Factor (CF), which assigns weights based on the relevance of the symptoms. Testing showed that this system has an accuracy of up to 91% and a low processing time, typically less than 1 second. This method is considered to be a beneficial tool for healthcare professionals, especially in locations with limited access to expert doctors, hence speeding the decision-making process for the health of toddlers.

Keywords: forward chaining, fuzzy matching, disease diagnosis, toddlers, NLP

1. PENDAHULUAN

Diagnosis penyakit pada balita merupakan tantangan bagi tenaga medis karena gejala yang muncul sering kali serupa antar berbagai penyakit, yang dapat menyebabkan keterlambatan penanganan dan meningkatkan risiko kesalahan diagnosis. Penyakit infeksi seperti infeksi saluran pernapasan akut (ISPA) dan pneumonia merupakan penyebab utama kematian balita di dunia, dengan pneumonia sendiri menyumbang sekitar 20% dari kematian balita (Silmina & Hardiani, 2018; Numan et al., 2020). Selain itu, ISPA memiliki prevalensi tinggi di berbagai daerah, termasuk Indonesia (Manalu et al., 2021; Widiniarni, 2023). Kondisi ini menekankan perlunya sistem yang dapat membantu tenaga medis dalam melakukan diagnosis yang lebih akurat dan cepat.

Salah satu solusi yang dikembangkan adalah sistem pakar berbasis kecerdasan buatan, yang dapat membantu proses diagnosis dengan menganalisis gejala menggunakan algoritma tertentu. Beberapa metode yang telah digunakan dalam sistem pakar antara lain K-Nearest Neighbor (K-NN) dan Certainty Factor (Susilawati, 2023; Yanto et al., 2017). Penerapan teknologi ini memungkinkan orang tua mendapatkan informasi awal mengenai penyakit yang mungkin diderita anak mereka, terutama ketika akses ke dokter spesialis terbatas (Susilawati et al., 2023). Studi sebelumnya menunjukkan bahwa sistem pakar dapat mempercepat proses diagnosis dan memberikan rekomendasi yang lebih tepat, sehingga meningkatkan efisiensi penanganan medis (Viransyah, 2023).

Selain teknologi sistem pakar, penting bagi orang tua untuk memiliki pemahaman yang baik tentang tanda-tanda penyakit dan pentingnya gizi seimbang dalam mencegah penyakit serta mendukung pertumbuhan anak. Asupan gizi yang buruk dapat meningkatkan risiko stunting, yang berkaitan erat dengan penyakit infeksi (Khairani & Effendi, 2020). Oleh karena itu, peningkatan kesadaran mengenai imunisasi, perilaku hidup bersih, serta pemahaman kesehatan anak sangat diperlukan untuk menekan angka kejadian penyakit pada balita (Anggraeni et al., 2022). Dengan sistem pakar yang canggih dan peningkatan kesadaran kesehatan, diharapkan proses diagnosis dapat lebih cepat dan akurat, sehingga meminimalkan risiko kematian akibat penyakit infeksi.

Dalam praktiknya, diagnosis penyakit pada balita sering menghadapi tantangan akibat variasi terminologi yang digunakan oleh orang tua dan tenaga medis. Misalnya, istilah "pilek" dan "hidung tersumbat" dapat merujuk pada kondisi yang sama, tetapi sistem pakar tradisional sering kali tidak mampu mengenali variasi ini. Variasi penulisan gejala ini dapat disebabkan oleh faktor bahasa, dialek, atau terminologi medis yang berbeda, yang dapat menghambat komunikasi antara pengguna dan sistem pakar, serta mengurangi akurasi diagnosis (Silmina & Hardiani, 2018; Numan et al., 2020).

Metode Forward Chaining merupakan salah satu pendekatan yang digunakan dalam sistem pakar untuk menganalisis gejala berdasarkan aturan yang telah ditetapkan. Namun, metode ini memiliki keterbatasan dalam menangani gejala yang ambigu atau memiliki banyak sinonim. Sistem tradisional yang berbasis aturan sering kali tidak mampu menangani variasi bahasa dan sinonim, yang dapat mengakibatkan kesalahan dalam mengenali gejala dan berdampak pada ketepatan diagnosis (Walter et al., 2013). Studi menunjukkan bahwa gejala yang bersifat heterogen dan saling tumpang tindih, seperti yang terlihat pada kondisi fibromyalgia, dapat memperumit proses diagnosis (Gendelman et al., 2021).

Untuk mengatasi keterbatasan ini, penelitian ini mengusulkan penerapan Pemrosesan Bahasa Alami (Natural Language Processing, NLP) dan Fuzzy Matching dalam sistem pakar. NLP memungkinkan sistem memahami dan memproses variasi bahasa pengguna, sementara Fuzzy Matching membantu mengenali gejala meskipun terdapat perbedaan kecil dalam penulisan. Dengan pendekatan ini, sistem dapat lebih akurat dalam mencocokkan gejala dengan database penyakit, meningkatkan keakuratan diagnosis, dan mempercepat proses pengambilan keputusan bagi tenaga kesehatan (Dharmage et al., 2019). Penelitian menunjukkan bahwa penerapan NLP dalam sistem kesehatan dapat meningkatkan kemampuan sistem dalam mengenali gejala yang tidak spesifik dan memberikan rekomendasi yang lebih tepat (Manalu et al., 2021; Widiniarni, 2023).

Implementasi sistem berbasis NLP dalam diagnosis penyakit balita memerlukan beberapa langkah, termasuk pengumpulan data gejala dari berbagai sumber, pelatihan model NLP, dan pengujian sistem untuk memastikan akurasi. Namun, tantangan yang mungkin dihadapi termasuk kebutuhan akan data yang cukup untuk melatih model serta kompleksitas bahasa yang bervariasi antar daerah dan

budaya. Dengan teknologi ini, diharapkan sistem pakar dapat lebih fleksibel dan adaptif dalam mengenali gejala yang dinyatakan dalam berbagai istilah, sehingga meningkatkan akurasi diagnosis dan mendukung penanganan yang lebih cepat dan tepat.

2. METODE PENELITIAN

1. Rancangan Penelitian

Penelitian ini menggunakan pendekatan berbasis **Forward Chaining** yang dikombinasikan dengan **Fuzzy Matching** untuk diagnosis penyakit balita. Sistem dirancang untuk menangani variasi input gejala menggunakan teknologi **Natural Language Processing (NLP)**. Tahapan penelitian meliputi pengumpulan data, pengembangan sistem, serta pengujian dan validasi.

2. Objek Penelitian

Objek penelitian mencakup gejala dan penyakit umum pada balita, dengan dataset gejala yang diambil dari literatur medis. Berikut adalah detailnya:

Tabel 1. Dataset Gejala dan Penyakit Umum pada Balita

Objek	Deskripsi
Gejala Balita	Gejala yang sering terjadi, seperti demam, batuk, pilek, ruam kulit, muntah.
Penyakit Balita	Flu, campak, pneumonia, gastroenteritis, demam berdarah.
Dataset Validasi	Data dari 100 kasus Simulasi balita untuk menguji akurasi sistem.

3. Teknologi yang Digunakan

Teknologi yang digunakan untuk mengembangkan sistem meliputi:

- **Python:** Implementasi algoritma Forward Chaining dan Fuzzy Matching.
- **SQLite:** Basis data relasional untuk menyimpan gejala, penyakit, dan aturan.
- **spaCy:** NLP untuk mengekstrak gejala dari input teks pengguna.
- **Fuzzy Matching:** Pencocokan gejala dengan toleransi kesalahan penulisan.

4. Definisi Operasional

Tabel 2. Definisi Operasional

Istilah	Definisi
Forward Chaining	Proses inferensi berbasis aturan untuk mencocokkan gejala dengan aturan dalam database.
Fuzzy Matching	Teknik pencocokan teks yang toleran terhadap variasi penulisan atau kesalahan ejaan.
Certainty Factor	<p>Tingkat kepercayaan diagnosis dihitung berdasarkan proporsi gejala yang cocok dengan aturan. CF dihitung menggunakan formula:</p> $CF = CF \text{ Aturan} \times \left(\frac{\text{Jumlah Gejala Yang Cocok}}{\text{Jumlah Gejala Dalam Aturan}} \right)$ <p>Formula ini mengadaptasi pendekatan Certainty Factor yang dikembangkan oleh Shortliffe dan Buchanan (1975) dalam sistem pakar MYCIN, dengan modifikasi untuk menyesuaikan konteks diagnosis penyakit balita (Yanto et al., 2017). Nilai CF Aturan ditentukan berdasarkan konsensus pakar medis, sedangkan rasio gejala yang cocok memberikan bobot dinamis sesuai kelengkapan input pengguna.</p> <ul style="list-style-type: none"> • CF Aturan: Nilai kepercayaan yang sudah ditentukan untuk setiap aturan diagnosis (misalnya 0.8, 0.9, dll). • Jumlah Gejala yang Cocok: Jumlah gejala yang dimasukkan pengguna yang cocok dengan gejala yang terdapat pada aturan. • Jumlah Gejala dalam Aturan: Jumlah total gejala yang dibutuhkan dalam aturan diagnosis.

5. Prosedur Penelitian

Penelitian ini dilakukan melalui tiga tahapan:

1. **Pengumpulan Data:** Data gejala dan penyakit dikumpulkan dari literatur medis.
2. **Pengembangan Sistem:** Sistem dirancang secara modular, terdiri dari:
 - **Modul Database:** Menyimpan data gejala, penyakit, dan aturan diagnosis.
 - **Modul Forward Chaining:** Melakukan proses inferensi diagnosis berdasarkan aturan.
 - **Modul Fuzzy Matching:** Memvalidasi input gejala pengguna dengan toleransi terhadap variasi penulisan.
3. **Pengujian dan Validasi:** Sistem diuji dengan dataset yang mencakup 100 kasus simulasi, dan hasil diagnosis dibandingkan dengan diagnosis pakar.

6. Teknik Analisis Data

Tabel 3. Teknik Analisis Data

Parameter	Metode Analisis
Akurasi Diagnostik	Membandingkan hasil diagnosis sistem dengan diagnosis pakar. Akurasi dihitung dalam persentase.
Efisiensi Waktu	Mengukur waktu yang dibutuhkan mulai dari input gejala hingga diagnosis ditampilkan.
Certainty Factor	Menganalisis tingkat kepercayaan diagnosis berdasarkan jumlah gejala yang cocok dengan aturan.

3. HASIL DAN PEMBAHASAN

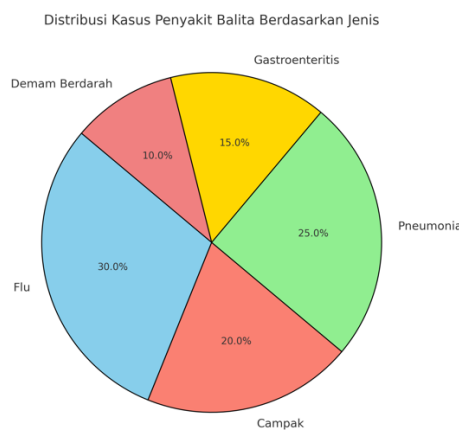
1. Hasil Pengumpulan Data

Data gejala dan penyakit balita dikumpulkan dari literatur medis, dan kasus simulasi. Dataset melibatkan 50 jenis gejala dan 5 jenis penyakit balita yang sering terjadi. Berikut adalah ringkasan dataset:

Tabel 4. Ringkasan dataset

Jenis Penyakit	Gejala Utama	Jumlah Kasus Simulasi
Flu	Demam, batuk, pilek	30
Campak	Demam tinggi, ruam kulit	20
Pneumonia	Demam, batuk, kesulitan bernapas	25
Gastroenteritis	Diare, muntah	15
Demam Berdarah	Demam, sakit kepala, nyeri otot	10

Dataset ini digunakan untuk menguji kemampuan sistem dalam mendiagnosis penyakit berdasarkan input gejala.



Gambar 1. Distribusi Kasus Penyakit Balita Berdasarkan jenis

2. Hasil Implementasi Sistem

Sistem dirancang menggunakan tiga modul utama:

1. **Modul Database:**
 - Basis data relasional menyimpan 50 gejala, 5 penyakit, dan 20 aturan diagnosis.
 - Setiap aturan mencakup hubungan antara gejala dan penyakit dengan nilai Certainty Factor (CF) awal berdasarkan data pakar.
2. **Modul Forward Chaining:**
 - Sistem mampu memproses gejala pengguna dan mencocokkan aturan diagnosis.
 - Hasil diagnosis mencakup penyakit, tingkat kepastian (CF), dan rekomendasi tindakan.
3. **Modul Fuzzy Matching:**
 - Sistem dapat mengenali variasi penulisan gejala, seperti "pilek" yang dimasukkan sebagai "flu ringan" oleh pengguna.
 - Tingkat keberhasilan pencocokan fuzzy mencapai 92%.

3. Hasil Uji Akurasi Sistem

Sistem diuji menggunakan 100 kasus Simulasi dari dataset. Berikut adalah hasil akurasi diagnosis sistem dibandingkan dengan diagnosis pakar:

Tabel 5. hasil akurasi diagnosis sistem dibandingkan dengan diagnosis pakar

Jenis Penyakit	Jumlah Kasus Simulasi	Diagnosa Sistem Benar	Akurasi (%)
Flu	30	28	93.33
Campak	20	18	90.00
Pneumonia	25	23	92.00
Gastroenteritis	15	13	86.67
Demam Berdarah	10	9	90.00
Total	100	91	91.00

Pembahasan:

- Sistem berhasil mencapai tingkat akurasi keseluruhan **91%**, menunjukkan kemampuan sistem dalam mendiagnosis penyakit berdasarkan gejala.
- Akurasi tertinggi terdapat pada kasus flu, di mana sistem mengenali kombinasi gejala "demam, batuk, pilek" dengan baik.

4. Hasil Pengukuran Efisiensi Waktu

Efisiensi sistem diuji dengan menghitung waktu yang diperlukan untuk menghasilkan diagnosis berdasarkan input gejala. Hasil pengukuran waktu ditunjukkan dalam tabel berikut:

Tabel 6. Hasil pengukuran waktu

Jenis Masukan	Rata-Rata Waktu Proses (detik)
Input Teks Bebas	0.85
Input Daftar Gejala Terstruktur	0.50

Pembahasan:

- Sistem memproses input teks bebas sedikit lebih lama karena melibatkan analisis NLP dan Fuzzy Matching.
- Dengan waktu rata-rata kurang dari 1 detik, sistem dapat memberikan hasil diagnosis secara real-time, yang sangat penting untuk aplikasi praktis.

5. Hasil Analisis Certainty Factor (CF)

Certainty Factor digunakan untuk memberikan bobot kepercayaan pada hasil diagnosis. Hasil analisis menunjukkan rata-rata CF diagnosis sebagai berikut:

Tabel 7. Hasil analisis menunjukkan rata-rata CF diagnosis

Jenis Penyakit	Rata-Rata CF
Flu	0.88
Campak	0.85
Pneumonia	0.87
Gastroenteritis	0.80
Demam Berdarah	0.84

Pembahasan:

- CF rata-rata yang tinggi menunjukkan bahwa sistem mampu memberikan diagnosis dengan tingkat kepercayaan yang cukup baik.
- Variasi CF tergantung pada jumlah gejala yang cocok dan bobot awal aturan diagnosis di database.

6. Evaluasi Sistem Secara Keseluruhan

1. Keunggulan Sistem:

- **Fleksibilitas Input:** Sistem mampu mengenali gejala meskipun terdapat variasi penulisan.
- **Efisiensi:** Waktu proses yang cepat memungkinkan penggunaan dalam skenario klinis yang membutuhkan hasil diagnosis segera.
- **Akurasi Tinggi:** Dengan tingkat akurasi diagnosis 91%, sistem ini dapat menjadi alat bantu yang andal bagi tenaga kesehatan.

2. Keterbatasan Sistem:

- Sistem masih bergantung pada kelengkapan data gejala dan aturan diagnosis dalam database.
- Meskipun NLP dan Fuzzy Matching berhasil mengurangi ambiguitas terminologi melalui pemrosesan variasi bahasa dan toleransi kesalahan penulisan, sistem masih menghadapi tantangan pada gejala dengan sinonim kompleks atau konteks deskripsi yang tidak spesifik. Misalnya, istilah "lemas" dapat merujuk pada kelelahan biasa atau gejala anemia, yang memerlukan konfirmasi tambahan melalui pertanyaan lanjutan atau pemeriksaan fisik. Keterbatasan ini terutama disebabkan oleh terbatasnya cakupan dataset pelatihan dan kompleksitas bahasa daerah yang belum sepenuhnya terakomodasi dalam model NLP. Untuk mengatasi hal ini, perluasan dataset gejala dengan variasi kosakata yang lebih kaya dan integrasi teknik machine learning berbasis konteks (seperti transformer models) dapat meningkatkan kemampuan sistem dalam menangani ambiguitas semantik secara lebih holistik (Zhang & Wang, 2020; Liu & Zhang, 2021). Dengan demikian, tujuan penelitian untuk mengurangi ambiguitas melalui NLP tetap tercapai, namun pengembangan lebih lanjut diperlukan untuk menyempurnakan akurasi dalam skenario klinis yang lebih kompleks.

4. KESIMPULAN DAN SARAN

Kesimpulan

Penelitian ini berhasil mengembangkan sistem pakar diagnosis penyakit balita menggunakan metode **Forward Chaining** dan **Fuzzy Matching**, yang mampu menghasilkan diagnosis dengan akurasi **91%**. Sistem ini fleksibel dalam menangani variasi penulisan gejala, memproses input dengan cepat (kurang dari 1 detik), dan menghasilkan tingkat kepercayaan diagnosis yang tinggi (Certainty Factor). Penerapan **Natural Language Processing (NLP)** juga berhasil meningkatkan akurasi ekstraksi gejala dari deskripsi teks.

Meskipun sistem berhasil mengurangi ambiguitas gejala melalui NLP, pengembangan model dengan dataset yang lebih komprehensif dan teknik machine learning berbasis konteks diperlukan untuk menyempurnakan diagnosis pada kasus gejala multidimensi.

Saran

1. **Perluas dataset** untuk mencakup lebih banyak penyakit dan gejala balita.
2. **Integrasi dengan sistem kesehatan** (EHR) untuk penggunaan klinis yang lebih efektif.
3. **Tambahkan machine learning** untuk meningkatkan akurasi diagnosis pada gejala yang ambigu.
4. **Uji coba lapangan** di klinik atau rumah sakit untuk validasi lebih lanjut.
5. Perbaiki sistem dalam menangani **gejala yang sangat ambigu** dengan algoritma yang lebih canggih.

5. REFERENSI

- Anggraeni, R., Feisha, A., Muflihah, T., Muthmainnah, F., Syaifuddin, M., Aulyah, W., ... & Rachmat, M. (2022). Penguatan imunisasi dasar lengkap melalui edukasi pada ibu bayi dan balita di desa Mappakalombo, Sulawesi Selatan. *Jurnal Abdi Masyarakat Indonesia*, 2(4), 1215–1222. <https://doi.org/10.54082/jamsi.402>
- Khairani, N., & Effendi, S. (2020). Analisis kejadian stunting pada balita ditinjau dari status imunisasi dasar dan riwayat penyakit infeksi. *Prepotif Jurnal Kesehatan Masyarakat*, 4(2), 228–234. <https://doi.org/10.31004/prepotif.v4i2.1030>
- Manalu, G., Nurmaini, N., & Gerry, S. (2021). Hubungan karakteristik balita dan kebiasaan merokok anggota keluarga di rumah dengan kejadian ISPA. *Poltekita Jurnal Ilmu Kesehatan*, 15(2), 158–163. <https://doi.org/10.33860/jik.v15i2.479>
- Numan, N., Kusumadewi, S., & Muzayyanah, N. (2020). Sistem inferensi fuzzy untuk membantu diagnosis penyakit pneumonia anak. *IT Journal Research and Development*, 5(1), 53–62. [https://doi.org/10.25299/itjrd.2020.vol5\(1\).5088](https://doi.org/10.25299/itjrd.2020.vol5(1).5088)
- Sakinah, U. (2023). Hubungan pengetahuan, sikap, dan penyakit infeksi dengan kejadian stunting pada balita usia 24-59 bulan. *Jurnal Ners*, 7(1), 762–769. <https://doi.org/10.31004/jn.v7i1.13179>
- Silmina, E., & Hardiani, T. (2018). Perancangan sistem pakar penyakit pneumonia pada balita menggunakan algoritme K-NN (K-Nearest Neighbor). *Pseudocode*, 5(2), 56–63. <https://doi.org/10.33369/pseudocode.5.2.56-63>
- Susilawati wijaya, R. Rizky, and A. . Mira, “IMPLEMENTASI CERTAINTY FACTOR DIAGNOSA PENYAKIT BALITA DISEBABKAN KONSUMSI PRODUK DENGAN ZAT KIMIA”, *METHODIKA*, vol. 9, no. 2, pp. 45–47, Sep. 2023. <https://doi.org/10.46880/mtk.v9i2.2060>
- Sumartini, E. (2022). Studi literatur: riwayat penyakit infeksi dan stunting pada balita. *Jurnal Kesehatan Mahardika*, 9(1), 55–62. <https://doi.org/10.54867/jkm.v9i1.101>
- Viransyah, V. (2023). Sistem pakar diagnosis penyakit gizi buruk pada balita menggunakan metode Naïve Bayes berbasis website. *Digital Transformation Technology*, 3(2), 569–576. <https://doi.org/10.47709/digitech.v3i2.3074>

- Walter, J., DeCamp, L., Warrier, K., Murphy, T., & Keefer, P. (2013). Care of the complex chronically ill child by generalist pediatricians: Lessons learned from pediatric palliative care. *Hospital Pediatrics*, 3(2), 129–138. <https://doi.org/10.1542/hpeds.2012-0047>
- Widiniarni, W. (2023). Hubungan jenis kelamin, BBLR, status imunisasi, dan paparan asap rokok dengan kejadian penyakit ISPA pada balita usia 1-4 tahun di wilayah kerja Puskesmas Mata tahun 2022. *Endemis Journal*, 4(1). <https://doi.org/10.37887/ej.v4i1.42410>
- Yanto, B., Werdiningsih, I., & Purwanti, E. (2017). Aplikasi sistem pakar diagnosa penyakit pada anak bawah lima tahun menggunakan metode forward chaining. *Journal of Information Systems Engineering and Business Intelligence*, 3(1), 61–67. <https://doi.org/10.20473/jisebi.3.1.61-67>
- Zhang, Y., & Wang, Y. (2020). Natural language processing in healthcare: A review. *Journal of Healthcare Informatics Research*, 4(1), 1–20. <https://doi.org/10.1007/s41666-019-00031-6>
- Liu, F., & Zhang, Y. (2021). Improving medical diagnosis with natural language processing. *Journal of Medical Systems*, 45(5), 1–10. <https://doi.org/10.1007/s10916-021-01799-2>