



Kombinasi *Modified K-Nearest Neighbor* dan *Certainty Factor* untuk Peningkatan Akurasi Diagnosis Malnutrisi

Zati Hulwani*, Eko Sedyono, Bayu Surarso

Magister Sistem Informasi, Sekolah Pascasarjana, Universitas Diponegoro

Naskah masuk: 21 Mei 2024; Diterima untuk publikasi: 3 Juli 2024
DOI: 10.21456/vol14iss3pp289-301

Abstract

Malnutrition remains a major public health challenge in low and middle-income countries. This research proposes a method for diagnosing malnutrition diseases by combining Certainty Factor (CF) and Modified K-Nearest Neighbor (MKNN). CF is used to obtain certainty values from the symptoms experienced by patients, while MKNN classifies patient symptom data into specific disease classes based on proximity to the training data. The symptom CF values are combined with the rule CFs to obtain the final CF for each disease. The patient's CF data becomes the testing data, while the disease dataset is the training data. The MKNN technique involves calculating the Euclidean distance, validity, and applying weight voting to identify the class to which the testing data belongs based on the majority class of the k nearest training data. In the test case, CF indicated a tendency towards Kwashiorkor, reinforced by MKNN with the majority of the nearest training data classified as Kwashiorkor. Cross-validation testing with 20 testing data resulted in an accuracy of 95% for the combined CF-MKNN method. The combination of the two methods mutually reinforces and increases the confidence in the diagnosis.

Keywords: Certainty Factor; Malnutrition; Modified K Nearest Neighbor; Expert System

Abstrak

Malnutrisi masih menjadi tantangan kesehatan utama di negara berpenghasilan rendah dan menengah. Penelitian ini mengusulkan metode diagnosis penyakit malnutrisi dengan menggabungkan *Certainty Factor* (CF) dan *Modified K-Nearest Neighbor* (MKNN). CF digunakan untuk mendapatkan nilai kepastian gejala pasien, sedangkan MKNN mengklasifikasikan data gejala ke dalam kelas penyakit berdasarkan kedekatan dengan data *training*. Nilai CF gejala dikombinasikan dengan CF aturan untuk memperoleh CF akhir setiap penyakit. MKNN menghitung jarak *Euclidean*, validitas, dan *weight voting* untuk menentukan kelas data *testing*. Pada kasus uji, CF mengindikasikan kecenderungan Kwashiorkor, dikuatkan MKNN dengan mayoritas data terdekat terklasifikasi Kwashiorkor. Pengujian *cross-validation* 20 data *testing* menghasilkan akurasi 95% metode gabungan CF-MKNN.

Kata kunci: *Certainty Factor*; Malnutrisi; *Modified K-Nearest Neighbor*; Sistem Pakar

1. Pendahuluan

Pada bab ini akan dijelaskan latar belakang pentingnya diagnosis malnutrisi secara dini dan akurat, serta gambaran umum mengenai metode yang diusulkan yaitu kombinasi *Certainty Factor* (CF) dan *Modified K-Nearest Neighbor* (MKNN) untuk mendiagnosis penyakit malnutrisi.

Malnutrisi tetap menjadi isu kesehatan masyarakat yang sangat berarti di seluruh dunia, terutama di negara-negara dengan penghasilan rendah dan menengah (Swinburn *et al.*, 2019). Kondisi ini tidak hanya menyebabkan masalah kesehatan akut seperti kekurangan energi protein, kekurangan vitamin dan mineral, tetapi juga berimplikasi jangka panjang berupa *stunting*, gangguan perkembangan kognitif,

penurunan produktivitas, dan peningkatan risiko penyakit kronis serta kematian dini (Victoria *et al.*, 2021). Upaya penanganan malnutrisi memerlukan pendekatan *multi* disiplin yang komprehensif, meliputi perbaikan akses pangan, sanitasi lingkungan, edukasi gizi, serta strategi berbasis teknologi untuk mengidentifikasi populasi berisiko secara akurat dan tepat sasaran.

Cara diagnosis tradisional yang hanya mengandalkan evaluasi klinis dan pemeriksaan fisik oleh tenaga medis sering kali kurang akurat dan berpotensi menyebabkan kesalahan dalam mengklasifikasikan kasus malnutrisi (Talukder and Ahammed, 2020). Oleh karena itu, diperlukan metode diagnosis yang lebih mutakhir dan terpercaya agar dapat mengidentifikasi malnutrisi dengan lebih tepat. Salah satu pendekatan yang menjanjikan adalah mengombinasikan metode *Certainty Factor* (CF) dan

*) Corresponding author: zatihulwaani@gmail.com

Modified K-Nearest Neighbor (MKNN) untuk memprediksi risiko malnutrisi.

MKNN merupakan pengembangan dari algoritma klasifikasi *K-Nearest Neighbors* (KNN) yang menggunakan jarak dan bobot keanggotaan untuk meningkatkan akurasi klasifikasi (Rojas *et al.*, 2019). Metode ini telah terbukti lebih akurat dibandingkan KNN konvensional dalam berbagai kasus klasifikasi, termasuk diagnosis penyakit (Ayyad *et al.*, 2019).

Sementara itu, CF merupakan metode untuk mengombinasikan bukti dari berbagai sumber dengan mengukur kepastian atau ketidakpastian suatu fakta atau aturan (Saputri *et al.*, 2020). Pendekatan ini memungkinkan penanganan data yang tidak sempurna atau tidak konsisten, yang sering terjadi dalam kasus malnutrisi dengan berbagai faktor penyebab yang kompleks.

Kombinasi MKNN dan CF diharapkan dapat memberikan hasil prediksi yang lebih akurat dan handal dalam mengidentifikasi individu atau komunitas berisiko malnutrisi. Dengan kemampuan prediksi yang lebih baik, sistem ini dapat membantu mengidentifikasi populasi berisiko sejak dini, sehingga intervensi dan pencegahan dapat dilakukan dengan lebih efektif dan tepat sasaran.

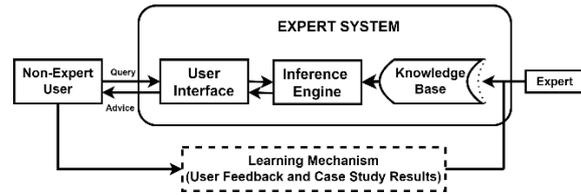
Pada penelitian ini akan dilakukan pada kasus penyakit malnutrisi dengan variabel yang digunakan adalah gejala-gejala dan jenis penyakit malnutrisi. Metode CF akan memberikan penilaian tingkat kepastian pada hasil klasifikasi yang dihasilkan oleh metode MKNN, yang kemudian akan menghasilkan diagnosis penyakit malnutrisi. Analisis kombinasi MKNN dengan CF diharapkan akan meningkatkan akurasi serta dapat memberikan informasi yang lebih baik bagi pasien dalam mengidentifikasi penyakit malnutrisi.

2. Kerangka Teori

Bab ini akan membahas landasan teori yang digunakan dalam penelitian, meliputi sistem pakar, metode CF, metode MKNN, dan penjelasan tentang penyakit malnutrisi.

2.1. Sistem Pakar (Expert System)

Sistem pakar adalah istilah yang digunakan untuk menyebut teknik cerdas yang sangat mengandalkan keahlian manusia, baik secara langsung maupun tidak langsung. Ketergantungan pada keahlian manusia secara langsung terjadi ketika ilmu dari para pakar tersebut diterapkan secara *online* ke dalam sistem (Halfon *et al.*, 2024). Sedangkan pengertian dari ketergantungan secara tidak langsung adalah ilmu dari para pakar akan diproses dan dirumuskan dalam bentuk peraturan-peraturan terlebih dahulu sebelum diintegrasikan ke dalam sistem. Berikut ini merupakan arsitektur dari sistem pakar yang dapat dilihat pada Gambar 1.



Gambar 1. Arsitektur Sistem Pakar

Berdasarkan Gambar 1. terlihat bahwa sistem pakar tersusun atas tiga komponen utama, yakni basis pengetahuan (*knowledge base*), mesin inferensi (*inference engine*), dan antarmuka pengguna (*user interface*). Basis pengetahuan berisi seluruh informasi, fakta, dan aturan-aturan yang dibutuhkan oleh sistem pakar untuk bekerja sesuai dengan rancangannya. Mesin inferensi menjalankan tindakan jika informasi yang diberikan oleh pengguna memenuhi ketentuan dalam aturan. Antarmuka pengguna menawarkan interaksi dengan pengguna non-ahli, di mana pengguna menjawab pertanyaan atau memasukkan data untuk memulai proses logis di mesin inferensi.

2.2. Modified K- Nearest Neighbor (MKNN)

MKNN adalah sebuah metode baru yang muncul sebagai hasil dari pengembangan yang dilakukan terhadap metode KNN. Sejumlah proses tambahan atau penyempurnaan diterapkan pada MKNN dibandingkan KNN. Salah satu alasan utama dilakukannya pengembangan dari KNN menjadi MKNN adalah untuk meningkatkan tingkat akurasi yang dimiliki KNN dengan cara menambahkan perhitungan nilai validitas. Hal ini didasari dugaan bahwa perhitungan bobot pada KNN memiliki kelemahan dalam menangani data yang memiliki nilai ekstrem (Stadnyk *et al.*, 2022). Adapun langkah-langkah yang terdapat dalam MKNN sebagai berikut:

1) Melakukan Perhitungan Jarak antar Data

Pendekatan MKNN merupakan metode klasifikasi yang mengelompokkan suatu objek ke dalam kelas tertentu berdasarkan kedekatan atau kedekatannya dengan objek-objek lain yang telah diketahui kelasnya sebelumnya. Penghitungan jarak antara titik pada data latih (x) dan titik pada data uji (y) dilakukan dengan mengaplikasikan *Euclidean* yang terdapat dalam Persamaan (1) (Bakthavachalam *et al.*, 2020).

$$D(x, y) = \sqrt{\sum_{k=1}^n (x_k - y_k)^2} \quad (1)$$

dengan

$D(x, y)$: Jarak *Euclidean* antara dua titik x dan y dalam ruang berdimensi n ;

$\sqrt{\quad}$: Simbol akar kuadrat;

Σ : Simbol sigma (Σ) yang mewakili operasi penjumlahan;

K : Indeks dimensi, dimulai dari 1 hingga n ;

$(x_k - y_k)^2$: Kuadrat dari selisih antara nilai pada dimensi ke-k dari titik x dan titik y;
 n : Jumlah dimensi dari ruang yang digunakan.

2) Melakukan Perhitungan Validitas Data Latih

Validitas dimanfaatkan untuk menghitung banyaknya data pada keseluruhan data latih, tetangga terdekat dari setiap data akan berpengaruh pada validitas data. Nilai validitas dihitung menggunakan Persamaan (2).

$$Validity = \frac{1}{K} \sum_{i=1}^K S(lbl(x), lbl(N_{i(x)})) \quad (2)$$

dengan

k : Jumlah total data atau sampel;
 Σ : Simbol sigma (Σ) yang mewakili operasi penjumlahan;
 I : Indeks atau penomoran data ke-i;
 $S(lbl(x), lbl(N_{i(x)}))$: Fungsi yang mengukur kesamaan (similarity) antara label prediksi $bl(x)$ dan label sebenarnya dari data $bl(N(x))$. Fungsi ini digunakan untuk menghitung kesamaan antara prediksi dan ground truth pada setiap sampel data.

Kemiripan antara titik x dengan data ke-i dari tetangga terdekatnya dihitung dengan menggunakan Fungsi S yang melibatkan Persamaan (3).

$$S(a, b) = \begin{cases} 1 & a=b \\ 0 & a \neq b \end{cases} \quad (3)$$

3) Perhitungan Bobot *Voting*

Penentuan bobot *voting* menggunakan validitas masing-masing data pada data latih dikalikan dengan bobot berdasarkan jarak *Euclidean*. Dalam metode MKNN, perhitungan bobot *voting* setiap tetangga dilakukan dengan Persamaan (4).

$$W(i) = Validitas(i) \times \frac{1}{d_{e+0,5}} \quad (4)$$

$W(i)$: Mewakili bobot (weight) pada waktu atau iterasi ke-i;
 Validitas(i) : Mewakili nilai validitas pada waktu atau iterasi ke-i;
 ∂e : Simbol turunan parsial.

Teknik ini memberi dampak berupa keuntungan yang lebih besar bagi sampel referensi yang memiliki validitas dan kedekatan yang lebih tinggi dengan

sampel uji. Keputusan yang dihasilkan tidak terlalu terpengaruh oleh sampel referensi yang memiliki kestabilan rendah di ruang fitur jika dibandingkan dengan sampel lainnya. Namun di sisi lain, mengalikan ukuran validitas dengan ukuran berbasis jarak dapat menjadi solusi atas kelemahan pembobotan berdasarkan jarak yang sering kali bermasalah dalam kasus data *outlier*. Dengan demikian, algoritma MKNN yang diajukan terbukti dapat lebih unggul dibandingkan metode KNN konvensional yang hanya mengandalkan perhitungan jarak semata.

MKNN memiliki sejumlah keunggulan dibanding KNN tradisional, di antaranya (Vahedifar et al., 2023):

- 1) Dapat memberikan bobot pada setiap fitur berdasarkan kontribusinya terhadap klasifikasi, tidak seperti KNN yang memperlakukan semua fitur sama rata.
- 2) Lebih efisien dalam menangani *dataset* besar karena hanya menggunakan sebagian data terdekat sebagai perwakilan, tidak seperti KNN yang menggunakan seluruh data latih .
- 3) Mampu beradaptasi dengan distribusi data yang tidak seimbang dan heterogen melalui pembobotan lokal pada data latih.

2.3. *Certainty Factor (CF)*

Certainty Factor (CF) adalah metode untuk mengukur tingkat kepastian pada sistem pakar. CF mengekspresikan derajat keyakinan dari suatu fakta dengan nilai antara 0 dan 1 (Sari et al., 2020). Formulasi dasar CF menggunakan Persamaan (5).

$$CF[H, E] = MB[H, E] - MD[H, E] \quad (5)$$

dengan:

H : Hipotesis;
 E : Fakta/Bukti;
 MB : Tingkat keyakinan (*Measure of Belief*);
 MD : Tingkat ketidakpercayaan (*Measure of Disbelief*).

Untuk mengombinasikan CF dari dua atau lebih aturan (R1 dan R2) yang menghasilkan kesimpulan yang sama, digunakan Persamaan (6).

$$CF(R1, R2) = CF(R1) + [CF(R2)] \times [1 - CF(R1)] \quad (6)$$

Selain itu, untuk kaidah dengan premis/gejala tunggal yang digunakan adalah Persamaan (7).

$$CF_{Keluhan}[H, E] = CF_{User}(H, E) \times CF_{Rules}(H, E) \quad (7)$$

Di samping itu, apabila terdapat kaidah dengan kesimpulan serupa atau melibatkan lebih dari satu gejala, maka penghitungannya menggunakan Persamaan (8).

$$CF_{combine} = CF_{fold} + CF_{gejala} \times (1 - CF_{fold}) \quad (8)$$

Lebih lanjut, guna mendapatkan persentase terhadap suatu penyakit, perhitungannya mengacu pada Persamaan (9).

$$CF_{\text{persentase}} = CF_{\text{combine}} \times 100\%. \quad (9)$$

Persamaan-persamaan tersebut digunakan untuk mengombinasikan dan menghitung *Certainty Factor* dalam sistem pakar.

2.4. Malnutrisi

Malnutrisi adalah istilah yang digunakan untuk merujuk pada penyakit yang disebabkan oleh kekurangan vitamin atau mineral dalam makanan. Lebih dari 2 miliar orang di dunia saat ini mungkin terkena malnutrisi. Malnutrisi utama di negara berkembang yaitu (Talukder and Ahammed, 2020):

1) Kekurangan Vitamin A

Kekurangan vitamin A, dalam beberapa kasus, menyebabkan lesi konjungtiva, *xerophthalmia* atau *keratomalaci*. Kondisi ini, jika tidak diobati, berkembang hingga penghancuran bilik mata depan, yang dapat menyebabkan kebutaan permanen. Pemberian vitamin A yang berlebihan secara kronis menyebabkan kerusakan hati, reversibel setelah penghentian suplementasi.

2) Kekurangan Zat Besi (*Fe*)

Kekurangan zat besi, otak menyebabkan berkurangnya sintesis *mielin* dan *neurotransmitter* dengan gangguan gerakan, memori dan kontrol persepsi. Kekurangan zat besi pada bayi dapat menyebabkan kerusakan otak yang *ireversibel*. Dalam selang waktu 6-12 bulan, kebutuhan *Fe* yang direkomendasikan untuk populasi adalah 11 mg/hari. Dari awal tahun kedua, menurun menjadi 7 mg/hari. Suplementasi hanya direkomendasikan dalam kasus defisiensi besi yang nyata.

3) Kekurangan Energi Protein

Malnutrisi energi protein dapat merusak beberapa jaringan tubuh manusia, seperti otak, sistem imun, dan mukosa usus. Di sisi lain, asupan protein yang berlebihan (> 15% E) tampaknya terkait dengan risiko kelebihan berat badan atau obesitas di masa dewasa, yang juga dapat meningkatkan risiko penyakit kronis.

4) Gangguan Akibat Kekurangan Yodium

Yodium sangat penting untuk sintesis hormon tiroid. Kekurangannya pada anak usia dini dapat menyebabkan gondok dan hipotiroidisme. Sebagian besar makanan kekurangan yodium, tetapi fortifikasinya memungkinkan mencapai tingkat asupan yang sesuai dengan usia, yaitu 90 g/hari.

2.5. Tinjauan Pustaka

Perkembangan teknologi pada *machine learning* dan *expert system* sudah diterapkan dalam berbagai bidang di kehidupan sehari-hari. Penelitian mengenai penggunaan algoritma *Modified K-Nearest Neighbor*

(MKNN), *Certainty Factor* (CF), dan malnutrisi sudah ada dilakukan oleh peneliti-peneliti sebelumnya. Berikut beberapa ulasan tentang penelitian-penelitian terdahulu yang relevan dengan penelitian ini.

Penelitian yang dilakukan oleh (Wafi *et al.*, 2019) mengenai prediksi prestasi akademik siswa dengan seleksi fitur otomatis. Penelitian ini menggunakan algoritma genetika untuk mencari kombinasi fitur terbaik dari sekumpulan fitur dengan MKNN yang digunakan untuk metode klasifikasi. Hasil dari penelitian ini menunjukkan bahwa seleksi fitur menggunakan algoritma genetika pada *Modified K-Nearest Neighbor* lebih baik dengan akurasi 82,6% dibandingkan dengan *K-Nearest Neighbor* yang hanya memiliki akurasi 73,6%. Hal tersebut membuktikan bahwa seleksi fitur dengan algoritma genetika bekerja dengan baik untuk menghilangkan fitur yang tidak berguna dan dapat meningkatkan metode klasifikasi.

Pada penelitian yang dilakukan oleh Cholissodin *et al.*, (2021) mengenai sistem klasifikasi tingkat laju data covid-19 untuk mitigasi penyebaran. Penelitian ini menggunakan metode *Modified K-Nearest Neighbor* (MKNN), di mana dalam metode ini, validasi dan pembobotan dilakukan dengan cara menghitung perbandingan antara tetangga yang berlabel sama dengan total keseluruhan tetangga. Faktor yang dijadikan parameter dalam proses klasifikasi adalah jumlah kasus positif, jumlah individu yang sembuh, dan jumlah korban jiwa akibat COVID-19. Penelitian ini menghasilkan sistem klasifikasi laju penyebaran COVID-19 dengan nilai akurasi tertinggi dari sistem yang diperoleh setelah dilakukan pengujian dengan nilai K sebanyak 3 dan α 0,5 adalah 97,794%.

Penelitian menggunakan metode CF juga dilakukan oleh (Sumiati *et al.*, 2021) dalam mengembangkan suatu sistem atau perangkat lunak yang dapat menggantikan seorang dokter untuk proses identifikasi kelainan jantung berdasarkan sistem pakar. Sistem pakar yang dikembangkan menggunakan metode CF dengan *pendekatan multiple rule*, dimana proses eksekusi aturan, beberapa aturan dapat menghasilkan suatu hipotesis. Pengujian sistem dilakukan dari hasil validitas sistem dengan pakar, sehingga hasil pengujian sistem menghasilkan nilai CF untuk jantung normal sebesar 0,95 dan tingkat akurasi 95%, sedangkan nilai *Certainty Factor* (CF) untuk abnormal jantung adalah 0,99 dan menghasilkan tingkat akurasi 99%.

Pada penelitian yang dilakukan oleh (Meidelfi *et al.*, 2021) mengkaji tentang sistem pakar untuk mendeteksi malnutrisi pada remaja dengan menggunakan teknik *forward chaining*. Dalam penelitian ini disebutkan bahwa pemahaman dan pengetahuan tentang gizi serta upaya perbaikan gizi sangat penting dalam melakukan pencegahan dari kemungkinan terburuk yang akan terjadi. Hasil dari penelitian ini berupa aplikasi sistem pakar berbasis

web sehingga dapat memudahkan pengguna untuk berkonsultasi di mana saja dan kapan saja. Meskipun demikian, sistem pakar ini masih dapat disempurnakan dengan memanfaatkan metode lain agar hasil penilaian yang diperoleh menjadi lebih optimal.

Penelitian yang dilakukan oleh Ibtasam (2021) dua metode digunakan yaitu *Naive Bayes* dan *Forward Chaining* untuk mengidentifikasi penyebab dan gejala malnutrisi. Metode *Naive Bayes* memberikan akurasi yang lebih tinggi yaitu 88% dibandingkan dengan *Forward Chaining* yang memiliki akurasi 85%. Kebaruan (*novelty*) yang diusulkan adalah menggabungkan pendekatan algoritma *Naive Bayes* dengan metode *Forward Chaining* untuk mendapatkan diagnosis malnutrisi yang akurat. Dengan menggabungkan kedua metode ini, diklaim dapat memberikan pengukuran penyebab malnutrisi yang sangat akurat.

Berdasarkan pada tinjauan pustaka yang telah dilakukan terlihat bahwa penggunaan metode KNN, MKNN, dan CF mampu mendiagnosis suatu penyakit dan studi kasus malnutrisi juga telah dilakukan dengan beberapa metode. Oleh karena itu, pada penelitian ini akan dilakukan:

- 1) Analisis kombinasi algoritma MKNN dengan CF pada kasus malnutrisi dengan variabel yang digunakan adalah gejala-gejala dan jenis penyakit malnutrisi.
- 2) Metode CF digunakan untuk memperoleh nilai kepercayaan terhadap penyakit yang telah ditentukan. Nilai kepercayaan diperoleh dari perhitungan metode CF dengan menggunakan nilai CF *user* dan CF pakar.
- 3) Algoritma MKNN digunakan untuk klasifikasi penyakit yang diderita oleh pasien berdasarkan gejala-gejala yang dialami. Hasil klasifikasi dari metode MKNN akan diberi nilai kepastian oleh metode CF sehingga menghasilkan suatu diagnosis penyakit malnutrisi.

3. Metode

Pada bab ini akan dijelaskan prosedur dan tahapan penelitian secara rinci, meliputi identifikasi masalah, studi literatur, pengumpulan data, perhitungan metode CF, perhitungan algoritma MKNN, dan penggabungan keduanya untuk mendapatkan hasil diagnosis penyakit malnutrisi.

3.1. Identifikasi Masalah

Pada tahap ini dilakukan identifikasi hal apa saja yang menjadi permasalahan di bidang kesehatan mengenai klasifikasi data dan *expert system* pada bidang kesehatan, sehingga dapat menentukan rencana kerja serta menentukan data apa saja yang dibutuhkan dalam penelitian ini. Pada penelitian ini, permasalahan yang muncul adalah apakah algoritma

MKNN dan CF dapat memecahkan masalah diagnosis penyakit malnutrisi.

3.2. Studi Literatur

Tahap ini dilakukan pencarian terhadap landasan-landasan teori terkait algoritma MKNN, CF, dan penyakit malnutrisi yang diperoleh dari berbagai sumber seperti buku, jurnal ilmiah dan juga referensi lainnya untuk melengkapi penelitian baik mengenai konsep dan teori sehingga memiliki acuan yang baik dan relevan.

3.3. Pengumpulan Data

Tahap ini melakukan studi dokumen dan studi lapangan yang terdiri dari observasi dan wawancara kepada pakar ahli gizi mengenai kasus penyakit malnutrisi, gejala-gejala, dan solusinya. Selain itu, dibutuhkan juga data pasien kasus malnutrisi yang didapatkan dari RSUD Drs. H. AMRI TAMBUNAN dari tahun 2000 - 2022. digunakan untuk membangun *dataset*. Disimpulkan dari data penyakit yang sering muncul pada *database* rekam medis RSUD Drs. H. AMRI TAMBUNAN, bahwa terdapat 3 penyakit yang sering muncul yaitu Kwashiorkor, Marasmus dan Marasmic-Kwashiorkor. Maka dari itu Representasi Pengetahuan yang dibangun pada sistem pakar dibatasi pada 3 penyakit tersebut dan 17 gejala yang telah ditentukan oleh pakar.

Setelah rekam medis terkumpul, peneliti melakukan data *preprocessing* untuk membentuk *dataset*. Data olah yang dimasukkan ke dalam *database* adalah nama samaran pasien, umur pasien, jenis kelamin pasien, gejala-gejala yang dialami pasien, dan hasil diagnosis dokter. Jumlah penyakit Kwashiorkor dalam *dataset* berjumlah 19 data, Marasmus 48 data dan Marasmic-Kwashiorkor berjumlah 33 data. Sehingga total keseluruhan *dataset* ada 100 data. Data yang tidak lengkap atau tidak konsisten akan dibuang untuk memastikan kualitas data dalam *dataset*.

Selain mengumpulkan dan membentuk *dataset*, diperlukan pembentukan Representasi Pengetahuan dalam membangun sistem pakar. Sistem pakar yang dibangun menerapkan model representasi Kaidah Produksi dan teknik inferensi *Forward Chaining*. Representasi pengetahuan dibentuk berdasarkan hasil wawancara dengan pakar ahli gizi mengenai gejala-gejala dan jenis penyakit malnutrisi, serta nilai CF yang diberikan oleh pakar untuk setiap aturan (*rule*) yang terkait, Dalam penjabarannya, berikut hasil representasi pengetahuan yang dibentuk untuk sistem pakar. dapat dilihat pada Tabel 1.

Tabel 1. Representasi pengetahuan

IF	Deskripsi Gejala	THEN	Nama Penyakit	CF Dokter
IF	Edema (pembengkakan)	THEN	Kwashiorkor	1
	Pandangan mata sayu			0,6

IF	Deskripsi Gejala	THEN	Nama Penyakit	CF Dokter	IF	Deskripsi Gejala	THEN	Nama Penyakit	CF Dokter
	Rambut tipis kemerahan			0,7		Kulit menjadi keriput			0,2
	Terjadi perubahan status mental menjadi apatis dan rewel			0,6		Jaringan lemak subkutis sangat sedikit sampai tidak ada			0,4
	Terjadi pembesaran hati			0,8		Perut cekung dan iga gambang			0,4
	Otot mengecil (hipotrofi)			0,9		Sering disertai penyakit infeksi			0,9
	Terdapat kelainan kulit berupa bercak merah			0,8		Diare kronik atau konstipasi (susah buang air)			0,8
	Sering disertai penyakit infeksi yang umumnya akut			0,6					
	Anemia dan diare			0,4					
IF	Badan nampak sangat kurus	THEN	Marasmus	1					
	Wajah seperti orang tua			0,9					
	Mudah menangis/cengeng dan rewel			0,6					
	Kulit menjadi keriput			0,8					
	Jaringan lemak subkutis sangat sedikit sampai tidak ada			0,8					
	Perut cekung dan iga gambang			1					
	Sering disertai penyakit infeksi			0,8					
	Diare kronik atau konstipasi (susah buang air)			0,2					
IF	Edema (pembengkakan)	THEN	Marasmic - kwashiorkor	0,6					
	Pandangan mata sayu			0,2					
	Rambut tipis kemerahan			0,4					
	Terjadi perubahan status mental menjadi apatis dan rewel			0,2					
	Terjadi pembesaran hati			0,2					
	Otot mengecil (hipotrofi)			0,4					
	Terdapat kelainan kulit berupa bercak merah muda			0,2					
	Sering disertai penyakit infeksi yang umumnya akut			0,9					
	Anemia dan diare			0,8					
	Badan nampak sangat kurus			0,6					
	Wajah seperti orang tua			0,4					
	Mudah menangis/cengeng dan rewel			0,9					

3.4. Perhitungan metode Certainty Factor

Perhitungan menggunakan metode *Certainty Factor* berfungsi untuk menangani ketidakpastian gejala. Proses yang dilakukan pada tahap ini adalah menetapkan Faktor Kepastian Aturan (*Rule Certainty Factor*) untuk setiap klasifikasi gejala yang ada. Penetapan aturan ini bertujuan untuk memberikan pembobotan kepastian pada masing-masing gejala dengan berpedoman pada basis aturan yang telah ditetapkan sebelumnya.. Selanjutnya, sistem menghitung CF dengan mengombinasikan CF dari pengguna (pasien) dan CF dari pakar (ahli gizi) untuk mendapatkan nilai CF gabungan. Kemudian, sistem menghitung CF *Combine* dengan mengaplikasikan aturan tertentu pada nilai CF gabungan. Hasilnya adalah nilai CF terbesar yang mewakili tingkat kepastian tertinggi dari gejala-gejala yang dialami pasien.

3.5. Perhitungan algoritma Modified K-Nearest Neighbor

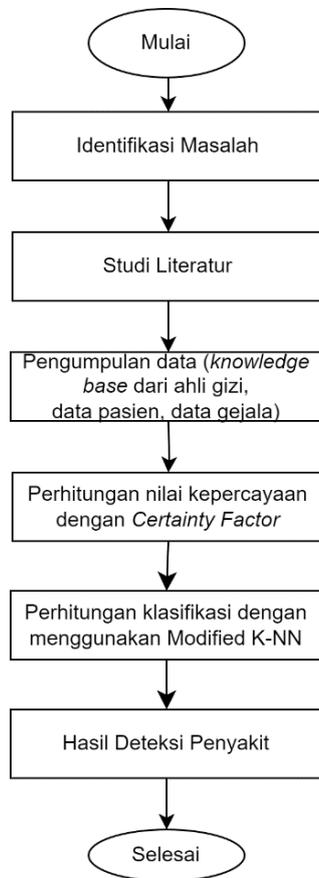
Perhitungan menggunakan metode MKNN yang berfungsi untuk mengklasifikasikan data pasien berdasarkan kemiripan dengan data latih. sistem mengeksekusi perhitungan jarak *Euclidean* antar data latih untuk mencari tetangga terdekat dari data uji (pasien). Setelah itu, sistem akan mengalkulasi nilai validitas untuk masing-masing data latih dengan mempertimbangkan kemiripan label kelas dari k tetangga terdekat yang telah diidentifikasi sebelumnya. Setelah itu, sistem akan melanjutkan dengan mengeksekusi kalkulasi jarak *Euclidean* antara data uji dengan setiap data latih yang ada, dengan tujuan untuk menemukan tetangga terdekat dari data uji tersebut. Kemudian, sistem melakukan *weight voting* dengan memberikan bobot tertinggi kepada tetangga terdekat dengan jarak *Euclidean* terkecil dan validitas tertinggi. Bobot ini dihitung dengan mengombinasikan validitas dan invers jarak *Euclidean*. Selanjutnya, sistem menentukan kelas data uji berdasarkan mayoritas kelas dari k tetangga terdekat dengan bobot voting tertinggi.

3.6. Hasil Diagnosis Penyakit

Hasil akhir dari proses ini adalah diagnosis penyakit malnutrisi beserta solusinya. Diagnosis

ditentukan berdasarkan kombinasi hasil dari metode CF yang menangani ketidakpastian gejala, dan metode MKNN yang mengklasifikasikan data pasien berdasarkan kemiripan dengan data latih melalui perhitungan jarak, validitas, dan *weight voting*.

Prosedur penelitian ditunjukkan dengan diagram alir pada Gambar 2. berikut:



Gambar 2. Diagram Alir Prosedur Penelitian

4. Hasil dan Pembahasan

Bab ini akan menyajikan hasil perhitungan dari metode yang diusulkan, baik metode CF maupun MKNN, serta pembahasan mengenai perbandingan dan penggabungan kedua metode tersebut dalam mendiagnosis penyakit malnutrisi.

4.1. Hasil

Berikut contoh data gejala dan interpretasi yang dirasakan pasien:

- 1) Edema (pembengkakan) – Mungkin Iya
- 2) Pandangan mata sayu – Mungkin tidak
- 3) Rambut tipis kemerahan dan mudah dicabut – Mungkin iya
- 4) Terjadi perubahan status mental menjadi apatis dan rewel – Mungkin tidak
- 5) Terjadi pembesaran hati – Mungkin iya
- 6) Otot mengecil (*hipotrofi*) – Mungkin iya
- 7) Terdapat kelainan kulit berupa bercak merah – Iya

- 8) Sering disertai penyakit infeksi yang umumnya akut – Mungkin tidak
- 9) Anemia dan diare – Mungkin tidak

Dari keluhan yang telah dimasukkan, interpretasi yang dirasakan pasien diubah menjadi nilai *Certainty Factor*. Sehingga detail keluhan yang dirasakan pasien adalah sebagai berikut:

- 1) Edema (pembengkakan) – 0.75
- 2) Pandangan mata sayu – 0.25
- 3) Rambut tipis kemerahan dan mudah dicabut – 0.75
- 4) Terjadi perubahan status mental menjadi apatis dan rewel – 0.25
- 5) Terjadi pembesaran hati – 0.75
- 6) Otot mengecil (*hipotrofi*) – 0.75
- 7) Terdapat kelainan kulit berupa bercak merah – 1
- 8) Sering disertai penyakit infeksi yang umumnya akut – 0.25
- 9) Anemia dan diare – 0.25

4.1.1. Diagnosis dengan Metode Certainty Factor

Untuk menjelaskan perhitungan dari diagnosis penyakit gigi menggunakan metode *certainty factor*, maka berikut simulasi perhitungan manualnya:

- 1) Dari contoh data gejala dan interpretasi yang dirasakan pasien di atas, keluhan-keluhan dipasangkan dengan penyakit menggunakan teknik inferensi. Karena terdapat 9 gejala, maka akan dicari seluruh aturan (*rules*) yang memiliki nilai IF dari gejala-gejala tersebut. Berikut rincian dari seluruh rules beserta nilai *CFRules* yang ada untuk contoh gejala tersebut:

- a. **IF** Edema **THEN** Kwashiorkor {CF: 1.0}
- b. **IF** Edema **THEN** Marasmic-Kwashiorkor {CF: 0.6}
- c. **IF** Pandangan mata sayu **THEN** Kwashiorkor {CF: 0.6}
- d. **IF** Pandangan mata sayu **THEN** Marasmic-Kwashiorkor {CF: 0.2}
- e. **IF** Rambut tipis kemerahan dan mudah dicabut **THEN** Kwashiorkor {CF: 0.7}
- f. **IF** Rambut tipis kemerahan dan mudah dicabut **THEN** Marasmic-Kwashiorkor {CF: 0.4}
- g. **IF** Terjadi perubahan status mental menjadi apatis dan rewel **THEN** Kwashiorkor {CF: 0.6}
- h. **IF** Terjadi perubahan status mental menjadi apatis dan rewel **THEN** Marasmic-Kwashiorkor {CF: 0.2}
- i. **IF** Terjadi pembesaran hati **THEN** Kwashiorkor {CF: 0.8}
- j. **IF** Terjadi pembesaran hati **THEN** Marasmic-Kwashiorkor {CF: 0.2}
- k. **IF** Otot mengecil (*hipotrofi*) **THEN** Kwashiorkor {CF: 0.9}
- l. **IF** Otot mengecil (*hipotrofi*) **THEN** Marasmic-Kwashiorkor {CF: 0.4}
- m. **IF** Terdapat kelainan kulit berupa bercak merah **THEN** Kwashiorkor {CF: 0.8}

- n. **IF** Terdapat kelainan kulit berupa bercak merah **THEN** Marasmic-Kwashiorkor {CF: 0.2}
 o. **IF** Sering disertai penyakit infeksi yang umumnya akut **THEN** Kwashiorkor {CF: 0.6}
 p. **IF** Sering disertai penyakit infeksi yang umumnya akut **THEN** Marasmic-Kwashiorkor {CF: 0.9}
 q. **IF** Anemia dan diare **THEN** Kwashiorkor {CF: 0.4}
 r. **IF** Anemia dan diare **THEN** Marasmic-Kwashiorkor {CF: 0.8}
- 2) Setelah menemukan seluruh aturan (rules), Selanjutnya setiap *CF rules* akan dikalikan dengan *CF Pasien* yang ada pada penjelasan di atas dengan menggunakan Persamaan (7):

- a. Perhitungan CF Keluhan Kwashiorkor
 Hasil dari perhitungan CF Keluhan pada penyakit Kwashiorkor dapat dilihat pada Tabel 2.

Tabel 2. Nilai CF Keluhan (Kwashiorkor)

CF	CF Rules	CF User	CF Keluhan
1	1	x	0.75
2	0.6	x	0.25
3	0.7	x	0.75
4	0.6	x	0.25
5	0.8	x	0.75
6	0.9	x	0.75
7	0.8	x	1
8	0.6	x	0.25
9	0.4	x	0.25

- b. Perhitungan CF Keluhan Marasmic-Kwashiorkor
 Hasil dari perhitungan CF Keluhan pada penyakit Marasmic-Kwashiorkor dapat dilihat pada Tabel 3.

Tabel 3. Nilai CF Keluhan (Marasmic-Kwashiorkor)

CF	CF Rules	CF User	CF Keluhan
1	0.6	x	0.75
2	0.2	x	0.25
3	0.4	x	0.75
4	0.2	x	0.25
5	0.2	x	0.75
6	0.4	x	0.75
7	0.2	x	1
8	0.9	x	0.25
9	0.8	x	0.25

- 3) Kemudian CF Keluhan dikelompokkan berdasarkan penyakit-penyakitnya masing-masing. Untuk penyakit yang memiliki nilai CF Keluhan lebih dari satu, maka Nilai CF antar

keluhan harus dikalkulasikan dengan perhitungan menggunakan Persamaan (8).

- a. Perhitungan *CFCombine* penyakit Kwashiorkor:

$$CFCombine(CF1,CF2) = CF1 + CF2 * (1-CF1) = 0.75 + 0.15 * (1-0.75) = 0.7875$$

$$CFCombine(CFold,CF3) = CFold + CF3*(1-CFold) = 0.7875 + 0.525*(1-0.7875) = 0.899$$

$$CFCombine(CFold,CF4) = CFold + CF4*(1-CFold) = 0.899 + 0.15*(1-0.899) = 0.9141$$

$$CFCombine(CFold,CF5) = CFold + CF5*(1-CFold) = 0.9141 + 0.6*(1-0.9141) = 0.9656$$

$$CFCombine(CFold,CF6) = CFold + CF6*(1-CFold) = 0.9656 + 0.675*(1-0.9656) = 0.988$$

$$CFCombine(CFold,CF7) = CFold + CF7*(1-CFold) = 0.988 + 0.8*(1-0.988) = 0.9984$$

$$CFCombine(CFold,CF8) = CFold + CF8*(1-CFold) = 0.9984 + 0.15*(1-0.9984) = 0.9986$$

$$CFCombine(CFold,CF9) = CFold + CF9*(1-CFold) = 0.9986 + 0.1*(1-0.9986) = 0.9987$$

Kesimpulan :

Dari perhitungan pertama yaitu perhitungan pada jenis penyakit kwashiorkor memiliki tingkat keyakinan sistem $0.9987 \approx 99,87\%$

- b. Perhitungan *CFCombine* penyakit Marasmic-Kwashiorkor:

$$CFCombine(CF1,CF2) = CF1 + CF2 * (1-CF1) = 0.45 + 0.05 * (1-0.45) = 0.4775$$

$$CFCombine(CFold,CF3) = CFold + CF3*(1-CFold) = 0.4775 + 0.3*(1-0.4775) = 0.6342$$

$$CFCombine(CFold,CF4) = CFold + CF4*(1-CFold) = 0.6342 + 0.05*(1-0.6342) = 0.6524$$

$$CFCombine(CFold,CF5) = CFold + CF5*(1-CFold) = 0.6524 + 0.15*(1-0.6524) = 0.7045$$

$$CFCombine(CFold,CF6) = CFold + CF6*(1-CFold) = 0.7045 + 0.3*(1-0.7045) = 0.79315$$

$$CFCombine(CFold,CF7) = CFold + CF7*(1-CFold) = 0.79315 + 0.2*(1-0.79315) = 0.8345$$

$$CFCombine(CFold,CF8) = CFold + CF8*(1-CFold) = 0.8345 + 0.225*(1-0.8345) = 0.8717$$

$$CFCombine(CFold,CF9) = CFold + CF9*(1-CFold) = 0.8717 + 0.2*(1-0.8717) = 0.89736$$

Kesimpulan :

Dari perhitungan kedua yaitu perhitungan pada jenis penyakit marasmic-kwashiorkor memiliki tingkat keyakinan sistem $0.8973 \approx 89,73\%$

- 4) Setelah menemukan semua nilai *Certainty Factor* setiap penyakit. Nilai *CFCombine* dibandingkan dan dicari yang terbesar
 - Kwashiorkor 99,87%
 - Marasmic – Kwashiorkor 89,73%
- 5) Dari dua perbandingan perhitungan di atas yaitu perhitungan pertama yaitu pada jenis penyakit kwashiorkor memiliki tingkat keyakinan sistem 99,87% dan pada perhitungan kedua jenis penyakit marasmic-kwashiorkor memiliki tingkat keyakinan sistem 89,73%. Maka dapat ditarik kesimpulan bahwa, pasien di diagnosa menderita penyakit kwashiorkor dengan tingkat keyakinan sistem 99.87%.

4.1.2. *Diagnosis dengan Metode Certainty Factor – Modified Nearest Neighbor*

- 1) Gejala dan interpretasi *Certainty Factor* yang dimasukkan pengguna akan dijadikan data *testing*. Sedangkan *dataset* akan dijadikan sebagai data *training*. Jika divisualisasikan ke dalam bentuk tabel, gejala yang dialami akan berisikan nilai *CF* dan sisa gejala yang tidak dialami akan bernilai 0, data *testing* akan terlihat seperti Tabel 4.

Tabel 4. Contoh data *testing*

GP1	GP2	GP3	..	GP14	GP15	GP16	GP17
0.75	0.25	0.75	..	0	0	0	0

- 2) Melakukan perhitungan metode MKNN dengan tahapan yaitu:
 - a. Sebagai langkah awal, proses yang harus dilakukan adalah normalisasi nilai data *training* dan data *testing* dengan menggunakan pendekatan normalisasi *min-max* dengan normalisasi persamaan (11).

$$normalisasi = \frac{CF\ gejala - \min(gejala)}{\max(gejala) - \min(gejala)} \quad (11)$$

Normalisasi : Mewakili proses normalisasi atau *penskalaan* nilai data

CF gejala : Mewakili nilai *Certainty Factor* (CF) dari suatu gejala atau atribut yang ingin dinormalisasi

min(gejala) : Mewakili nilai minimum dari keseluruhan data gejala yang ada

max(gejala) : Mewakili nilai maksimum dari keseluruhan data gejala yang ada

Karena menghitung normalisasi memerlukan nilai *minimum* (terkecil) dan nilai *maximum* (terbesar) dari setiap gejala. Maka seluruh *dataset* dan data *testing* digabungkan dan dicari nilai terbesar dan terkecilnya per gejala. Pada contoh kali ini, nilai *minimum* dan *maximum* per gejala seperti terlihat pada Tabel 5.

Tabel 5. Nilai minimum dan maximum tiap gejala

Min/Max	GP 1	GP 2	GP 3	..	GP 14	GP 15	GP 16	GP 17
Min	0	0	0	..	0	0	0	0
Max	1	1	1	..	1	1	1	1

Normalisasi data testing:

Diketahui nilai CF Pasien terhadap gejala yang dialaminya adalah $CF_{Pasien}[GP01] = 0.75$, $CF_{Pasien}[GP02] = 0.25$, $CF_{Pasien}[GP03] = 0.75$, $CF_{Pasien}[GP04] = 0.25$, $CF_{Pasien}[GP05] = 0.75$, $CF_{Pasien}[GP06] = 0.75$, $CF_{Pasien}[GP07] = 1$, $CF_{Pasien}[GP08] = 0.25$. $CF_{Pasien}[GP09] = 0.25$.

$$normalisasi [GP01] = \frac{0.75-0}{1-0} = 0.75$$

$$normalisasi [GP02] = \frac{0.25-0}{1-0} = 0.25$$

$$normalisasi [GP03] = \frac{0.75-0}{1-0} = 0.75$$

$$normalisasi [GP04] = \frac{0.25-0}{1-0} = 0.25$$

$$normalisasi [GP05] = \frac{0.75-0}{1-0} = 0.75$$

$$normalisasi [GP06] = \frac{0.75-0}{1-0} = 0.75$$

$$normalisasi [GP07] = \frac{1-0}{1-0} = 1$$

$$normalisasi [GP08] = \frac{0.25-0}{1-0} = 0.25$$

$$normalisasi [GP09] = \frac{0.25-0}{1-0} = 0.25$$

Untuk nilai normalisasi gejala lain pasti akan bernilai 0, sebab pembilang dari pembagian normalisasi bernilai 0 (angka 0 jika dibagi bilangan apa pun akan menghasilkan nilai 0).

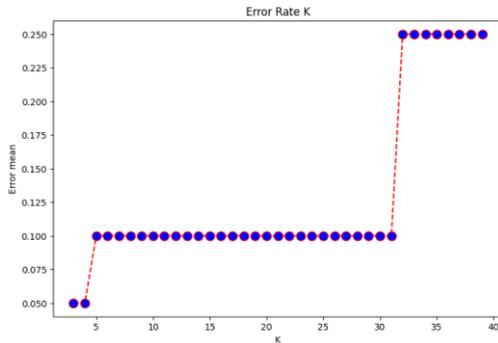
Sama seperti data testing, setiap *CFPasien* yang tidak bernilai 0 akan memiliki nilai normalisasi 0 karena pembilang dari pembagian normalisasi bernilai 0 (angka 0 jika dibagi bilangan apa pun akan menghasilkan nilai 0). Hasil normalisasi data *training* jika divisualisasikan ke dalam sebuah tabel dapat dilihat pada Tabel 6.

Tabel 6. Data *training* ternormalisasi

Nama	GP 1	GP 2	GP 3	GP 4	..	GP 15	GP 16	GP 17
Pasien 1	1	0.75	0.75	1	..	0	0	0
Pasien 2	0.50	0	0	0	..	0	1	1
Pasien 32	0.75	0.25	0.75	0.25	..	0	0	0
..
Pasien 100	0.75	0.25	0.50	0	..	0.5	1	0.75

- b. Menentukan nilai k
 Proses berikutnya adalah penentuan nilai k. Dalam pelaksanaan ini, nilai k yang ditentukan adalah k=3. Pemilihan nilai k berdasarkan

perhitungan laju *error* (*error rate*) pada Gambar 3.



Gambar 3. Grafik *Error Rate K*

Berdasarkan gambar, nilai *k* yang optimal adalah *k*=3. Beberapa alasan yang mendukung pemilihan nilai *k*=3 sebagai nilai terbaik berdasarkan grafik tersebut:

- *Error rate* terendah
 Pada grafik, terlihat bahwa *error rate* terendah dicapai pada nilai *k*=3. Titik ini menunjukkan performa terbaik model dalam hal meminimalkan kesalahan prediksi.
 - Stabilitas *error rate*
 Untuk nilai *k* kecil, yakni *k*=2 dan *k*=3, *error rate* tampak relatif stabil dan datar. Ini mengindikasikan bahwa model tidak terlalu sensitif terhadap perubahan nilai *k* di sekitar 2-3. Stabilitas ini menguntungkan dalam pemilihan nilai *k*.
 - Titik belok grafik
 Setelah *k*=3, grafik *error rate* mulai menunjukkan kenaikan yang cukup curam. Titik belok ini terjadi sekitar *k*=10. Nilai *k*=3 berada tepat sebelum titik belok, menunjukkan bahwa peningkatan *k* lebih lanjut akan menyebabkan penurunan performa yang signifikan.
 - *Trade-off bias-variance*
 Nilai *k* yang lebih rendah seperti 3 cenderung mengurangi varians model tetapi meningkatkan bias. Meski demikian, bias rendah lebih disukai daripada varians tinggi yang dapat menyebabkan *overfitting*.
- c. Mengeksekusi perhitungan jarak Euclidean antara data-data latih
- Proses selanjutnya adalah melakukan kalkulasi jarak *Euclidean* antara data-data latih. Tahapan ini bertujuan untuk menemukan jarak antara data latih yang digunakan dengan mengaplikasikan perhitungan sesuai dengan persamaan (1). Berikut Langkah-langkah untuk mencari nilai jarak Euclidean:

- Perhitungan data latih 1 dan 2

$$d(1,2) = \sqrt{(1-0.5)^2 + (0.75-0.5)^2 + (0.75-0)^2 + (1-0)^2 + (0.25-0)^2 + (1-0.5)^2 + (0.75-0)^2 + (0.5-1)^2 + (0.25-0.75)^2 + (0-0.25)^2 + (0-0)^2 + (0-0.75)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-1)^2 + (0-1)^2}$$

$$d(1,2) = 2.524876$$

- Perhitungan data latih 1 dan 3

$$d(1,3) = \sqrt{(1-0.25)^2 + (0.75-0)^2 + (0.75-0)^2 + (1-0)^2 + (0.25-0)^2 + (1-0)^2 + (0.75-0)^2 + (0.5-0.75)^2 + (0.25-1)^2 + (0-0.25)^2 + (0-0)^2 + (0-1)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-1)^2 + (0-0.75)^2}$$

$$d(1,3) = 2.75$$

Perhitungan di atas dilakukan pada semua data latih untuk memperoleh jarak *euclidean*, perhitungan untuk data latih yang lain menggunakan persamaan yang sama dan perhitungan yang sama, nilai *euclidean* yang diperoleh dapat dilihat pada Tabel 7.

Tabel 7. Hasil perhitungan jarak euclidean antar data-data latih

Data ke	1	2	3	4	..	98	99	100
1	0.00	2.524	2.750	3.072	..	3.020	3.041	2.345
2	2.524	0.00	0.750	2.384	..	2.449	2.397	1.118
..
100	2.345	1.118	1.299	1.920	..	1.968	1.968	0

- d. Mengeksekusi perhitungan nilai validitas
- Untuk mendapatkan nilai validitas menggunakan hasil dari perhitungan jarak *Euclidean* dengan mengambil rentang terdekat sebanyak nilai *k*, karena *k* yang digunakan ada 5 maka dari Tabel 7. diambil 5 nilai terkecil dari setiap data latih, setelah itu ditentukan apakah masing-masing 3 nilai terkecil itu merupakan kelas yang sama, jika kelas sama maka nilainya 1 jika tidak sama maka bernilai 0, menghitung validitas menggunakan persamaan (2).

$$\text{Validitas}(1) = \frac{1}{3} \sum_{i=1}^3 s(\text{label}(x), (\text{label}(Ni(x)))) = \frac{1}{3} (1+1+1) = 1$$

$$\text{Validitas}(2) = \frac{1}{3} \sum_{i=1}^3 s(\text{label}(x), (\text{label}(Ni(x)))) = \frac{1}{3} (1+1+1) = 1$$

Nilai validitas dilakukan kepada semua data yang ada sehingga menemukan data 1 dan 0, data yang dihasilkan dapat dilihat pada Tabel 8.

Tabel 8. Hasil perhitungan validitas data *training*

Data latih ke	K=1	K=2	K=3	Sum(a,b)	Validitas
1	1	1	1	3	1
2	1	1	1	3	1
3	1	1	1	3	1
..
63	1	1	1	3	1
..
84	1	1	1	3	1

e. Mengeksekusi perhitungan jarak Euclidean pada data uji

Selanjutnya adalah menemukan nilai *Euclidean* dengan melibatkan data uji. Berikut adalah cara yang diterapkan untuk menghitung jarak antara data latih dan data uji.

- Perhitungan data uji dan data latih 1

$$d(uji, latih1) = \sqrt{(0.75 - 1)^2 + (0.25 - 0.75)^2 + (0.75 - 0.75)^2 + (0.25 - 1)^2 + (0.75 - 0.25)^2 + (0.75 - 1)^2 + (1 - 0.75)^2 + (0.25 - 0.5)^2 + (0.25 - 0.25)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2}$$

$$d(uji, latih1) = 1.145644$$

- Perhitungan data uji dan data latih 99

$$d(uji, latih99) = \sqrt{(0.75 - 0)^2 + (0.25 - 0)^2 + (0.75 - 0)^2 + (0.25 - 0)^2 + (0.75 - 0)^2 + (0.75 - 0)^2 + (1 - 0)^2 + (0.25 - 0)^2 + (0.25 - 0)^2 + (0 - 0.75)^2 + (0 - 1)^2 + (0 - 0.5)^2 + (0 - 0.75)^2 + (0 - 0.5)^2 + (0 - 0.75)^2 + (0 - 1)^2 + (0 - 0)^2}$$

$$d(uji, latih 99) = 2.772634$$

Kemudian dilakukan perhitungan yang sama untuk semua data uji, hasil perhitungan *euclidian* data uji 1 ditunjukkan pada Tabel 9.

Tabel 9. Hasil perhitungan Euclidean data uji dan data latih

Data ke-	Euclidean data latih dan data uji
1	1.145644
2	2.410913
7	0.707107
8	0.661438
26	2.291288
31	0.559017
64	0.750000
65	0.790569
67	0.612372
97	2.549510
99	2.772634

f. Mengeksekusi penghitungan *weight voting*.

Proses ini merupakan tahapan terakhir dari klasifikasi *Modified K-Nearest Neighbor*. Pada tahap ini, kelas akan ditentukan oleh nilai *weight voting* tertinggi yang diperoleh. Berikut langkah menghitung *weight voting* menggunakan Persamaan (4).

$$W(1) = 1 \times \frac{1}{1.145644 + 0.5} = 0.60765$$

$$W(7) = 1 \times \frac{1}{0.707107 + 0.5} = 0.82842$$

Selanjutnya melakukan perhitungan dengan metode yang serupa untuk semua data uji yang tersedia, sementara hasil dari penghitungan *weight voting* untuk data uji 1 ditampilkan pada Tabel 10.

Tabel 10. Hasil perhitungan *weight training*

Data ke-	Weight Voting	Diagnosis
1	0.60765	Kwashiorkor
2	0.34353	Marasmic-kwashiorkor
7	0.82842	Kwashiorkor
8	0.86100	Kwashiorkor
26	0.13714	Marasmic-kwashiorkor
31	0.94427	Kwashiorkor
64	0.80000	Kwashiorkor
65	0.77485	Kwashiorkor
67	0.89897	Kwashiorkor
97	0.32792	Marasmus
99	0.12865	Marasmus

g. Melakukan penentuan kelas dari data uji

Setelah mendapatkan nilai *weight voting* dari seluruh data latih, maka langkah selanjutnya adalah mencari nilai *weight voting* terbesar. Nilai tertinggi dari data yang terdekat akan dijadikan sebagai dasar untuk memperoleh hasil klasifikasi. Pengelompokan mayoritas nilai tertinggi terdekat akan dilakukan berdasarkan nilai k yang ditetapkan, yaitu k = 3, sehingga 3 data dengan nilai tertinggi dari tetangga terdekat akan menjadi dasar untuk menentukan kelas. Nilai tertinggi dari *weight voting* dapat dilihat pada Tabel 11.

Tabel 11. Pencarian tiga nilai tertinggi dari *weight voting*

K	Weight Voting	Diagnosis
K1	0.94427	Kwashiorkor
K2	0.89897	Kwashiorkor
K3	0.86100	Kwashiorkor
Mayoritas		Kwashiorkor(benar)

Hasil *voting* pada tabel di atas menunjukkan bahwa data testing di diagnosa sebagai penyakit **Kwashiorkor**.

4.1.3. Akurasi Metode Certainty Factor – Modified K Nearest Neighbor

Metode yang digunakan adalah *Cross Validation* 20 data testing (dari *dataset*) yang sama digunakan untuk metode yang lain agar mendapat kesamaan yang serupa, sehingga pengujian tingkat akurasi dengan menggunakan skala perbandingan 20:80 ditampilkan pada Tabel 12.

Tabel 12. *Cross Validation*

	Marasmus	Marasmic-Kwashiorkor	Kwashiorkor
Marasmus	3	1	0
Marasmic-Kwashiorkor	0	5	0
Kwashiorkor	0	0	11

$$Accuracy = \frac{3 + 5 + 11}{20} = 0.95$$

Akurasi model klasifikasi dengan 3 kelas: Marasmus, Marasmic-Kwashiorkor, dan Kwashiorkor dihitung dari jumlah diagonal utama dalam tabel *confusion matrix* dibagi total jumlah data, sehingga menghasilkan akurasi 0,95 atau 95%.

Tabel *confusion matrix* ini menunjukkan bahwa dari 20 data testing, 3 data diklasifikasikan dengan benar sebagai Marasmus, 5 data diklasifikasikan dengan benar sebagai Marasmic-Kwashiorkor, dan 11 data diklasifikasikan dengan benar sebagai Kwashiorkor. Terdapat 1 data Marasmus yang salah diklasifikasikan sebagai Marasmic-Kwashiorkor.

Metode *cross validation* yang digunakan adalah dengan membagi data menjadi 20 bagian, kemudian melatih model dengan 19 bagian dan menguji pada 1 bagian sisanya, secara bergantian hingga semua bagian data digunakan sebagai data testing. Hal ini membantu mengevaluasi kinerja model secara lebih komprehensif.

Dari hasil perhitungan di atas dihasilkan bahwa akurasi dari pengujian diagnosis penyakit malnutrisi menggunakan metode *Certainty Factor – Modified K Nearest Neighbor* adalah sebesar 95%.

4.2. Pembahasan

Penelitian ini mengusulkan metode diagnosis penyakit malnutrisi dengan mengombinasikan metode *Certainty Factor* dan *Modified K-Nearest Neighbor*. Metode *Certainty Factor* digunakan untuk mendapatkan nilai kepastian atau *certainty factor* dari gejala-gejala yang dialami pasien terhadap kemungkinan penyakit yang diderita. Sedangkan metode MKNN digunakan untuk mengklasifikasikan data testing (data gejala pasien) ke dalam kelas penyakit tertentu berdasarkan kedekatan dengan data *training* (data penyakit).

Langkah awal adalah menghitung nilai CF untuk setiap kemungkinan penyakit berdasarkan gejala yang dialami pasien. Gejala-gejala diinterpretasikan terlebih dahulu ke dalam nilai CF mulai dari 0 sampai 1. Kemudian nilai CF gejala dikombinasikan dengan

nilai CF aturan (*rules*) menggunakan persamaan tertentu untuk mendapatkan nilai CF akhir untuk setiap penyakit. Dari perhitungan ini, diperoleh nilai CF tertinggi untuk penyakit Kwashiorkor sebesar 99,87%.

Untuk memvalidasi hasil diagnosis menggunakan CF, digunakan metode MKNN. Data CF pasien yang berisi nilai CF untuk setiap gejala dijadikan data testing, sedangkan *dataset* penyakit malnutrisi yang berisi nilai CF untuk setiap gejala dan kelas penyakitnya menjadi data *training*. Data *testing* dan data *training* dinormalisasi terlebih dahulu menggunakan *min-max normalization* agar berada dalam rentang 0 sampai 1.

Selanjutnya dihitung jarak *Euclidean* antara data *testing* dengan setiap data *training*. Berdasarkan jarak ini, dihitung nilai validitas untuk setiap data *training* menggunakan fungsi validitas tertentu. Lalu, *weight voting* dihitung dengan mengombinasikan nilai validitas dan jarak *Euclidean* antara data pengujian terhadap setiap data pelatihan menggunakan persamaan *weight voting*.

Pada tahap akhir klasifikasi MKNN, k data *training* dengan nilai *weight voting* tertinggi yang terdekat dengan data testing diambil untuk menentukan kelas data testing. Pada kasus ini, k=3, sehingga 3 data *training* dengan *weight voting* tertinggi diambil. Mayoritas kelas dari 3 data ini akan menjadi kelas untuk data testing. Hasilnya, mayoritas 3 *weight voting* tertinggi adalah untuk kelas kwashiorkor, sama dengan hasil diagnosis menggunakan metode CF sebelumnya.

Pada pengujian akurasi, dilakukan *cross validation* dengan 20 data testing dari *dataset* yang sama. Hasil pengujian menunjukkan akurasi metode gabungan CF-MKNN sebesar 95%. Nilai akurasi yang tinggi ini mengindikasikan metode gabungan ini cukup baik digunakan untuk diagnosis penyakit malnutrisi berdasarkan gejala. Kombinasi dua metode CF dan MKNN memberikan hasil diagnosis yang saling memperkuat dan meningkatkan kepercayaan terhadap keputusan diagnosis akhir.

5. Kesimpulan

Pengujian *cross-validation* dengan 20 data testing menghasilkan akurasi 95% dalam diagnosis malnutrisi berdasarkan data uji. Akurasi ini lebih tinggi dibandingkan metode diagnosis malnutrisi sebelumnya menggunakan metode *Naïve Bayes* dengan akurasi 88% dan *Forward Chaining* yang memiliki akurasi 85%. Hal ini menunjukkan kombinasi CF dan MKNN mampu meningkatkan akurasi dan keandalan diagnosis.

Daftar Pustaka

Ayyad, S.M., Saleh, A.I., Labib, L.M., 2019. Gene Expression Cancer Classification using Modified

- K-Nearest Neighbors Technique. *Biosystems*, 176, 41-51. <https://doi.org/10.1016/j.biosystems.2018.12.009>
- Bakthavachalam, M.D., Albert S., Raj, A., 2020. A Study on Breast Cancer Analysis by using K-Nearest Neighbor with Different Distances and Classification Rules using Machine Learning. *European Journal of Molecular and Clinical Medicine*, 7(3), 4842-4851.
- Cholissodin, I., Evanita, F.M., Tedjasulaksana, J.J., Wahyuditomo, K.W., 2021. Klasifikasi Tingkat Laju Data Covid-19 untuk Mitigasi Penyebaran Menggunakan Metode Modified K-Nearest Neighbor (MKNN). *Jurnal Teknologi Informasi dan Ilmu Komputer*, 8(3), 595-600. <https://doi.org/10.25126/jtiik.2021834400>
- Halfon, P., Penaranda, G., Ringwald, D., Retornaz, F., Boissel, N., Bodard, S., Feryn, J.M., Bensoussan, D., Cacoub, P., 2024. Laboratory Tests for Investigating Anemia: From an Expert System to Artificial Intelligence. *Practical Laboratory Medicine*, 39, e00357. <https://doi.org/10.1016/j.plabm.2024.e00357>
- Ibtasam, M., 2021. Accuracy Measurements and Decision Making by Na⁺-ve Bayes and Forward Chaining Method to Identify the Malnutrition Causes and Symptoms. *Scientific Journal of Informatics*, 8(2), 320-324. <https://doi.org/10.15294/sji.v8i2.29317>
- Meidelfi, D., Alfarissy, S., Fauzi, A., Azura, R., 2021. Sistem Pakar Mendeteksi Malnutrisi pada Remaja dengan Metode Forward Chaining. *Jurnal Teknologi Informasi dan Komunikasi*, 11(1), 25. <http://dx.doi.org/10.35585%2Finspir.v11i1.2608>
- Rojas, J.S., Rendon, A., Corrales, J.C., 2019. Consumption Behavior Analysis of Over the Top Services: Incremental Learning or Traditional Methods?. *IEEE Access*, 7, 136581-136591. <https://doi.org/10.1109/ACCESS.2019.2942782>
- Saputri, A.E., Sevani, N., Saputra, F., Sali, R.K., 2020. Using Certainty Factor Method to Handle Uncertain Condition in Hepatitis Diagnosis. *ComTech: Computer, Mathematics and Engineering Applications*, 11(1), 1-10. <https://doi.org/10.21512/comtech.v11i1.5903>
- Sari, R.D., Rahmadani, R., Putri, T.T.A., Hutahaean, H.D., Sriadhi, S., 2020. Certainty Factor and Dempster-Shafer Method Analysis for Early Detection of Final Year Student Depression. *IOP Conference Series: Materials Science and Engineering*, 830(3), 1-7. <https://doi.org/10.1088/1757-899X/830/3/032007>
- Stadnyk, M., Fryz, M., Zagorodna, N., Muzh, V., Kochan, R., Nikodem, J., Hamera, L., 2022. Steady State Visual Evoked Potential Classification by Modified KNN Method. *Procedia Computer Science*, 207, 71-79. <https://doi.org/10.1016/j.procs.2022.09.039>
- Sumiati, Saragih, H., Rahman, T.K.A., Triayudi, A., 2021. Expert System for Heart Disease Based on Electrocardiogram Data using Certainty Factor with Multiple Rule. *IAES International Journal of Artificial Intelligence*, 10(1), 43-50. <https://doi.org/10.11591/ijai.v10.i1.pp43-50>
- Swinburn, B.A., et al., 2019. The Global Syndemic of Obesity, Undernutrition, and Climate Change: The Lancet Commission report. *The Lancet*, 393 (10173), 791-846. [https://doi.org/10.1016/S0140-6736\(18\)32822-8](https://doi.org/10.1016/S0140-6736(18)32822-8)
- Talukder, A., Ahammed, B., 2020. Machine Learning Algorithms for Predicting Malnutrition Among Under-Five Children in Bangladesh. *Nutrition*, 78, 110861. <https://doi.org/10.1016/j.nut.2020.110861>
- Vahedifar, M. A., Akhtarshenas, A., Sabbaghian, M., Rafatpanah, M. M., Toosi, R., 2023. *Information Modified K-Nearest Neighbor*. <http://arxiv.org/abs/2312.01991>
- Victora, C.G., Christian, P., Vdaletti, L.P., Gatica-Domínguez, G., Menon, P., Black, R.E., 2021. Revisiting Maternal and Child Undernutrition in Low-Income and Middle-Income Countries: Variable Progress Towards an Unfinished Agenda. *The Lancet*, 397(10282), 1388-1399. [https://doi.org/10.1016/S0140-6736\(21\)00394-9](https://doi.org/10.1016/S0140-6736(21)00394-9)
- Wafi, M., Faruq, U., Supianto, A.A., 2019. Automatic Feature Selection for Modified K-Nearest Neighbor to Predict Student's Academic Performance. *IEEE*. <https://doi.org/10.1109/SIET48054.2019.8986074>