

Sistem Diagnosa Dini *Menstrual Disorders* Dengan K-NN

Diterima: 10 Juni 2024
Revisi: 10 Juli 2024
Terbit: 1 Agustus 2024

**¹Yuniswatin Nahdiyah, ²Daniel Swanjaya, M.Kom, ³Risky Aswi
Ramadhani, ⁴Septian Geges**
¹⁻³Universitas Nusantara PGRI Kediri, ⁴Universitas Palangka Raya
*¹yuniswatin123@gmail.com, ²swanjayadaniel@gmail.com,
³risky_aswi@unpkediri.ac.id, ⁴septian.geges@it.upr.ac.id*

Abstrak—Pada setiap bulannya seorang perempuan yang sedang tidak hamil dan juga *menopause* mengalami menstruasi. Menstruasi menjadi tanda perkembangan sistem reproduksi perempuan. Pada saat menstruasi kebanyakan perempuan mengalami gejala-gejala dari gangguan menstruasi baik secara fisik maupun psikis. Gangguan menstruasi tersebut tentunya dapat mempengaruhi aktivitas hariannya dan tidak dapat diabaikan. Pada penelitian ini dibuat sistem diagnosa dini *menstrual disorders* dengan menggunakan metode k-nn, dengan tujuan dapat membantu para perempuan dalam melakukan diagnosa tanpa perlu ke dokter. Pada pembuatan sistem dilakukan pengujian dengan membandingkan nilai k agar didapatkan hasil yang optimal. Pada penelitian ini dengan dilakukan uji *t-test* didapatkan model k-nn dengan k=3 dan k=7 mampu memberikan hasil yang baik berdasarkan ukuran data pelatihan. Terdapat perbedaan signifikan secara statistik antara performa model knn dengan nilai k=1 dan k=7 yang menghasilkan *P-Value* kurang dari 0.05.

Kata Kunci—Sistem pakar; K-Nearest Neighbor; Uji T

Abstract— Every month a woman who is not pregnant and also menopausal experiences menstruation. Menstruation is a sign of the development of the female reproductive system. During menstruation, most women experience symptoms of menstrual disorders both physically and psychologically. Menstrual disorders can certainly affect daily activities and cannot be ignored. In this research, an early diagnosis system for menstrual disorders was created using the k-nn method, with the aim of helping women make a diagnosis without needing to see a doctor. When creating the system, testing is carried out by comparing the k values to obtain optimal results. In this research, by carrying out the t-test, it was found that the k-nn model with k=3 and k=7 was able to provide good results based on the size of the training data. There is a statistically significant difference between the performance of the knn model with values k=1 and k=7 which produces a P-Value of less than 0.05.

Keywords—Expert System; K-Nearest Neighbor; T-Test

This is an open access article under the CC BY-SA License.



Penulis Korespondensi:

Daniel Swanjaya, M.Kom,
Universitas Nusantara PGRI Kediri,
Email : swanjayadaniel@gmail.com

I. PENDAHULUAN

Seseorang pada saat mengalami masa transisi dari remaja menjadi dewasa atau yang biasa disebut dengan pubertas mengalami perubahan dari segi fisik, biologis, psikis maupun sosial [1]. Pada seorang perempuan saat mengalami pubertas ditandai dengan mulainya menstruasi yang menandakan perkembangan reproduksinya [2]. Seorang perempuan yang mengalami menstruasi biasanya berusia 12-49 tahun yang tidak sedang hamil ataupun sudah *menopause* [3]. Pada saat menstruasi para perempuan biasanya mengalami beberapa gangguan baik secara fisik maupun psikis. Gangguan menstruasi merupakan kelainan yang berhubungan dengan perubahan yang terjadi pada masa menstruasi seperti perubahan siklus, jumlah darah dan lainnya [4]. Namun kebanyakan perempuan mengabaikan masalah tersebut dan tidak mengetahui dampak dari gangguan menstruasi. Hal tersebut dikarenakan kurangnya informasi dan pengetahuan mengenai kesehatan reproduksi bagi perempuan. Selain itu terkadang para perempuan malu untuk berkonsultasi ke dokter. Pada perkembangan teknologi, bidang kedokteran memanfaatkannya dengan meningkatkan pelayanan dalam mendiagnosa penyakit [5]. Maka diperlukan sebuah sistem yang dapat digunakan tanpa perlu pergi ke dokter atau seorang pakar agar lebih mudah [6].

Sistem pakar merupakan sistem yang dimana dalam melakukan pemecahan masalah didasarkan pada pengetahuan manusia yang dimasukkan kedalam komputer [7]. Dengan perkembangan teknologi dan informasi yang semakin maju dapat dimanfaatkan dalam pembuatan sistem pakar. Web dapat digunakan untuk mempermudah pengguna dalam melakukan interaksi secara langsung melalui internet dengan penyedia informasi atau pakar [8]. Sehingga para perempuan dapat melakukan diagnosa gangguan menstruasi menggunakan sistem pakar. Dalam mendiagnosa gangguan menstruasi digunakan metode yaitu *K-Nearest Neighbor* (KNN). *K-Nearest Neighbor* adalah algoritma yang ada pada data mining dan digunakan untuk proses klasifikasi [9]. *K-Nearest Neighbor* mengklasifikasikan data dengan mencari jarak terdekat antara data yang dievaluasi dengan k tetangga terdekat dari data latih [10]. Namun pemilihan parameter k yang terlalu besar atau kecil dapat memengaruhi performa knn [11]. Pada KNN untuk menghindari munculnya jarak yang sama pada saat proses klasifikasi, nilai k umumnya dibuat ganjil (3,5,7) [12]. Untuk itu diperlukan evaluasi performa KNN dan melakukan pembetulan parameter k untuk memperoleh nilai k yang optimal [13]. Berdasarkan uraian latar belakang diatas maka tujuan dari penelitian ini adalah mengetahui akurasi dari penerapan metode *K-Nearest Neighbor* dengan menggunakan uji *T-Test* dengan penggunaan nilai k dan pembagian data yang berbeda.

II. METODE

Dalam penelitian ini akan digunakan metode penelitian dan pengembangan yang dimana hasil dan juga keefektifan sistem dapat digunakan dan bermanfaat bagi masyarakat. Metode pengumpulan data dilakukan dengan studi literatur, observasi dan wawancara.

2.1 *K-Nearest Neighbor* (K-NN)

Berdasarkan atribut dan juga *sample* pelatihan metode *K-Nearest Neighbor* mampu mengklasifikasikan objek baru berdasarkan jarak terdekat. Perhitungan jarak terdekat menggunakan rumus *Euclidean distance* [14]. Dimana pada proses perhitungan KNN terdapat langkah sebagai berikut :

1. Menentukan parameter k
2. Menghitung jarak menggunakan *Euclidean distance*, seperti berikut

$$dis(x_1, x_2) = \sqrt{\sum_{i=0}^n (x_{1i} - x_{2i})^2}$$

Gambar 1. Rumus *Euclidean distance*

Dimana :

x_1 = data uji

x_2 = Sampel data

i = Variabel data

dis = jarak

n = dimensi data

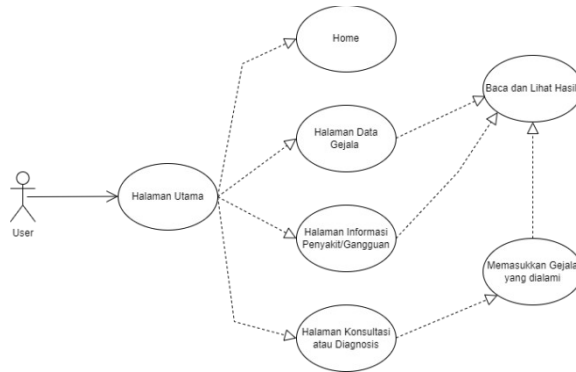
3. Mengurutkan jarak (urutan naik)
4. Menentukan jarak terdekat
5. Mencari jumlah kelas dari tetangga yang terdekat.

2.2 Uji *T-Test*

T-Test merupakan metode pengujian hipotesis yang menggunakan satu objek penelitian dengan menggunakan dua perlakuan yang berbeda [15]. Meskipun pada metode *T-Test* menggunakan objek penelitian yang sama, namun sampel data tetap terbagi menjadi dua yaitu data dengan perlakuan yang pertama juga data dengan perlakuan yang kedua. Dengan *T-Test* dapat diketahui signifikan atau tidaknya metode K-NN.

2.3 Desain Sistem

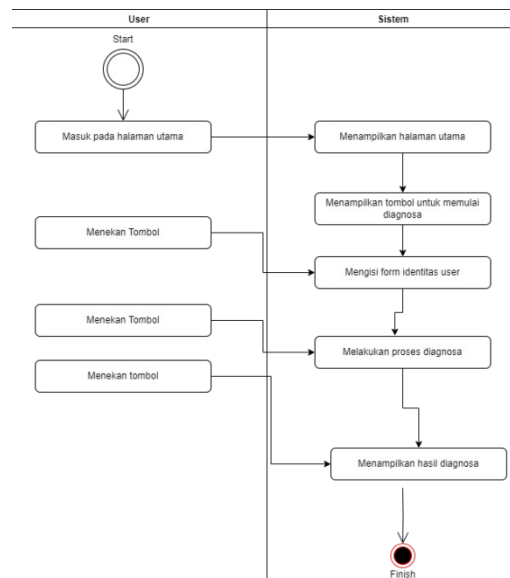
a. Use case diagram



Gambar 2. Use case diagram

Pada gambar 2 merupakan *use case* dari sistem diagnosa dini *menstrual disorders*, dimana *user* dapat melakukan proses diagnosa dan juga melihat informasi mengenai data gejala dan juga data penyakit atau gangguan yang ada pada masa menstruasi.

b. Activity diagram



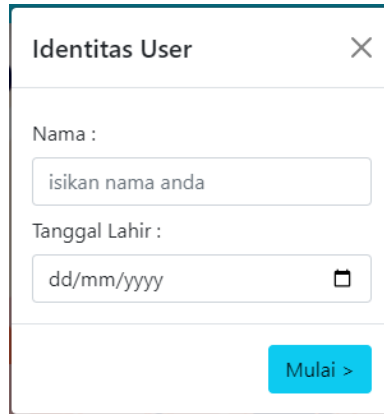
Gambar 3. Activity diagram

Pada gambar 3 adalah *activity diagram* dari sistem. Dimana alur dari sistem dimulai dari *user* masuk ke dalam sistem, kemudian menekan tombol 'Mulai Diagnosa'. Setelah itu *user* diminta untuk mengisi *form* identitas *user*. Kemudian *user* dapat menekan tombol 'Mulai' dan dapat melakukan proses diagnosa dengan menginputkan jawaban berdasar pertanyaan. Setelah itu menekan tombol *submit* dan *user* dapat mengetahui hasil dari proses diagnosa yang telah dilakukan.

III. HASIL DAN PEMBAHASAN

3.1 Implementasi Hasil

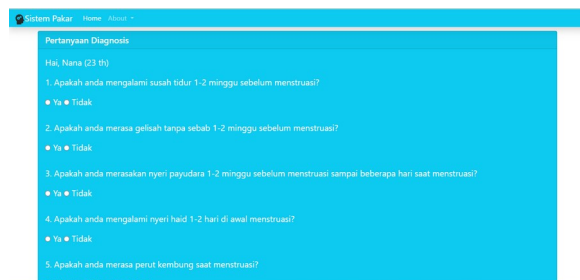
a. *Form identitas user*



Gambar 4. *Form identitas user*

Pada gambar 4 merupakan *form identitas user*. Pada *form* ini *user* harus menginputkan nama dan juga tanggal lahir agar dapat melakukan proses diagnosa.

b. **Halaman diagnosa**



Gambar 5. **Halaman diagnosa**

Jenis Penyakit	Persentase
Dismenore	40.00%
Menometroragia	0.00%
Menorhagia	0.00%
PMDD	0.00%
PMS	20.00%
Sehat	20.00%

Hasil Diagnosa: Dismenorea (40.00%)

[Kembali](#)

Gambar 6. **Halaman hasil**

Pada gambar 5 merupakan halaman diagnosa. Pada halaman tersebut *user* diminta untuk menginputkan jawaban berdasarkan gejala yang dialami. kemudian *user* menekan tombol

submit dan akan dilakukan proses klasifikasi. Sehingga akan didapatkan hasil diagnosa yang akan ditampilkan pada halaman hasil diagnosa seperti pada gambar 6.

3.2 Pengujian K-NN

Pengujian metode k-nn dengan *t-test* dilakukan menggunakan perbandingan ukuran data pelatihan 60%, 70% dan 80%.

```
Train size: 60.0%
Cross-validated accuracies for k=1: [0.13888889 0.22222222 0.33333333]
Cross-validated accuracies for k=3: [0.30555556 0.25 0.30555556]
Cross-validated accuracies for k=5: [0.27777778 0.25 0.25 ]
Cross-validated accuracies for k=7: [0.25 0.25 0.30555556]
T-Test between k=1 and k=3: T-Statistic = -0.9370425713316365, P-Value = 0.4017876631161809
T-Test between k=1 and k=5: T-Statistic = -0.48666426339228736, P-Value = 0.6519586963751551
T-Test between k=1 and k=7: T-Statistic = -0.6246950475544247, P-Value = 0.5660371777000405
T-Test between k=3 and k=5: T-Statistic = 1.3416407864998738, P-Value = 0.25081535976844566
T-Test between k=3 and k=7: T-Statistic = 0.7071067811865461, P-Value = 0.5185185185185193
T-Test between k=5 and k=7: T-Statistic = -0.4472135954999597, P-Value = 0.6778688286992482
Test accuracy for k=1: 0.1917808219178082
Test accuracy for k=3: 0.1917808219178082
Test accuracy for k=5: 0.2054794520547945
Test accuracy for k=7: 0.2465753424657534
```

Gambar 7. *T-Test* dengan ukuran data pelatihan 60%

Pada gambar 7 merupakan hasil dari pengujian dengan ukuran data pelatihan 60%, dimana didapatkan akurasi terbaik dengan parameter k=7. Tetapi tidak ada perbedaan yang signifikan secara statistik antara performa model dengan berbagai nilai k dikarenakan pada pengujian tersebut semua nilai dari *P-Value* lebih besar dari 0.05 (semua *p-value* >0.05).

```
Train size: 70.0%
Cross-validated accuracies for k=1: [0.24 0.28 0.2 ]
Cross-validated accuracies for k=3: [0.28 0.36 0.24]
Cross-validated accuracies for k=5: [0.2 0.32 0.28]
Cross-validated accuracies for k=7: [0.36 0.4 0.32]
T-Test between k=1 and k=3: T-Statistic = -1.2649110640673522, P-Value = 0.2745766290948479
T-Test between k=1 and k=5: T-Statistic = -0.632455532033676, P-Value = 0.5614380442505256
T-Test between k=1 and k=7: T-Statistic = -3.6742346141747677, P-Value = 0.021311641128756713
T-Test between k=3 and k=5: T-Statistic = 0.5345224838248489, P-Value = 0.6213082950374968
T-Test between k=3 and k=7: T-Statistic = -1.5811388300841909, P-Value = 0.18900365845517508
T-Test between k=5 and k=7: T-Statistic = -2.2135943621178664, P-Value = 0.09126024444398564
Test accuracy for k=1: 0.30303030303030304
Test accuracy for k=3: 0.15151515151515152
Test accuracy for k=5: 0.2727272727272727
Test accuracy for k=7: 0.18181818181818182
```

Gambar 8. *T-Test* dengan ukuran data pelatihan 70%

Pada gambar 8 merupakan hasil dari pengujian dengan ukuran data pelatihan 70% yang didapatkan akurasi terbaik pada model k=1. Pada pengujian ini terdapat perbedaan yang signifikan secara statistik antara performa model dengan k=1 dan k=7. Karena pada pengujian dengan parameter k=1 dan k=7 didapatkan nilai dari *P-Value* yang nilainya lebih kecil dari 0.05 (*P-Value* < 0.05).

```
Train size: 80.0%
Cross-validated accuracies for k=1: [0.3 0.2 0.2]
Cross-validated accuracies for k=3: [0.3 0.15 0.2 ]
Cross-validated accuracies for k=5: [0.3 0.3 0.3]
Cross-validated accuracies for k=7: [0.2 0.25 0.3 ]
T-Test between k=1 and k=3: T-Statistic = 0.3015113445777636, P-Value = 0.778049548900021
T-Test between k=1 and k=5: T-Statistic = -2.000000000000001, P-Value = 0.11611652351681541
T-Test between k=1 and k=7: T-Statistic = -0.37796447300922786, P-Value = 0.7246586364960951
T-Test between k=3 and k=5: T-Statistic = -1.8898223650461365, P-Value = 0.13177756749413527
T-Test between k=3 and k=7: T-Statistic = -0.6324555320336763, P-Value = 0.5614380442505255
T-Test between k=5 and k=7: T-Statistic = 1.7320508075688774, P-Value = 0.15830242337545783
Test accuracy for k=1: 0.2
Test accuracy for k=3: 0.3333333333333333
Test accuracy for k=5: 0.2666666666666666
Test accuracy for k=7: 0.2
```

Gambar 9. *T-Test* dengan ukuran data pelatihan 80%

Pada gambar 9 merupakan hasil dari pengujian dengan ukuran data pelatihan 80% yang didapatkan hasil akurasi terbaik dengan nilai $k=3$. Pada pengujian ini tidak ada perbedaan yang signifikan secara statistik antara performa model dengan berbagai nilai k , karena semua nilai dari *P-Value* lebih besar dari 0.05 ($P\text{-Value} > 0.05$).

IV. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan pada sistem diagnosa dini *menstrual disorders* didapatkan hasil kesimpulan yaitu dengan melakukan peningkatan ukuran pada data pelatihan mampu menghasilkan hasil yang lebih stabil pada nilai k yang lebih tinggi. Pada model knn dengan nilai $k=3$ atau $k=7$ mampu memberikan hasil yang baik berdasarkan ukuran data pelatihan, dan juga pada ukuran data pelatihan 70% dengan $k=1$ dan $k=7$ terdapat perbedaan yang signifikan secara statistik antara performa model knn dengan nilai k yang mampu menghasilkan *P-Value* lebih kecil dari 0.05.

DAFTAR PUSTAKA

- [1] A. Wahyuni and Y. R. Ichianti, "Deteksi Dini Gangguan Haid Dalam Upaya Peningkatan Kesehatan Reproduksi Remaja," *JMM (Jurnal Masy. Mandiri)*, vol. 7, no. 4, p. 3428, 2023, doi: 10.31764/jmm.v7i4.15908.
- [2] M. Minanti and N. Y. S. Munti, "Sistem Pakar Diagnosa Gejala Sindrom Premenstruasi (Pms) Menggunakan Metode Certainty Factor Berbasis Website," *Skanika*, vol. 5, no. 1, pp. 60–71, 2022, doi: 10.36080/skanika.v5i1.2919.
- [3] I. Novia, dan Nunik Puspitasari, R. Kabupaten Sidoarjo, and D. Biostatistika dan Kependudukan Fakultas Kesehatan Masyarakat Universitas Airlangga, "Faktor Risiko yang Mempengaruhi Kejadian Dismenore Primer".
- [4] D. Andreswari, A. Erlansari, F. Farady Coastera, and J. Pebrianty Hasian Lumbanraja,

“Implementasi Metode Certainty Factor Dalam Sistem Pakar Diagnosis Awal Tanda Bahaya Gangguan Menstruasi,” *J. Rekursif*, vol. 11, no. November, pp. 70–84, 2023, [Online]. Available: www.ejournal.unib.ac.id/index.php/rekursif

- [5] A. N. D. Soetarmono, “Perancangan Sistem Pakar Dalam Mendiagnosa Penyakit Pada Balita,” *Teknika*, vol. 2, no. 1, pp. 28–39, 2013, doi: 10.34148/teknika.v2i1.11.
- [6] A. Rasheda and T. Arifin, “Penerapan K-Nearest Neighbor Untuk Sistem Pakar Diagnosa Penyakit Tulang Belakang,” vol. 3, no. 2, 2022.
- [7] R. Helilintar and R. A. Ramadhani, “Sistem Pakar Diagnosis Hepatitis Menggunakan Metode K-NN untuk Pelayanan Kesehatan Primer,” *S N a T I K a*, vol. 4, pp. 19–23, 2017.
- [8] S. Boll, “Media Informatics,” *IT - Inf. Technol.*, vol. 51, no. 6, pp. 301–302, 2009, doi: 10.1524/itit.2009.9059.
- [9] F. L. D. Cahyanti, W. Gata, and F. Sarasati, “Implementasi Algoritma Naïve Bayes dan K-Nearest Neighbor Dalam Menentukan Tingkat Keberhasilan Immunotherapy Untuk Pengobatan Penyakit Kanker Kulit,” *J. Ilm. Univ. Batanghari Jambi*, vol. 21, no. 1, p. 259, 2021, doi: 10.33087/jiubj.v21i1.1189.
- [10] T. A. Munandar and A. Q. Munir, “Implementasi K-Nearest Neighbor Untuk Prototype Sistem Pakar Identifikasi Dini Penyakit Jantung,” *Respati*, vol. 17, no. 2, p. 44, 2022, doi: 10.35842/jtir.v17i2.457.
- [11] R. T. Prasetio, “SELEKSI FITUR DAN OPTIMASI PARAMETER k-NN BERBASIS ALGORITMA GENETIKA PADA DATASET MEDIS,” *J. Responsif Ris. Sains dan Inform.*, vol. 2, no. 2, pp. 213–221, 2020, doi: 10.51977/jti.v2i2.319.
- [12] T. Nurhidayat, P. Kasih, and A. Sanjaya, “Classification of Aglaonema Plants Berdasarkan Corak Daun,” *Semin. Nas. Inov. Teknol.*, pp. 223–228, 2019, [Online]. Available: <https://proceeding.unpkediri.ac.id/index.php/inotek/article/view/541>
- [13] A. M. Wibowo, P. Kasih, and I. N. Farida, “Sistem Bantu Penentuan Konsentrasi Mahasiswa Menggunakan Metode K-Nearest Neighbor Classification,” *Semin. Nas. Teknol. Sains*, vol. 3, no. 1, pp. 370–379, 2024, doi: 10.29407/stains.v3i1.4343.
- [14] M. F. H. Siregar, I. N. Farida, and D. M. A. Widyadara, “Penerapan Metode SAW Dan KNN Untuk Deteksi Dini Kanker Serviks,” *Inotek*, vol. 7, pp. 2549–7952, 2023.
- [15] Ardiyansyah, P. A. Rahayuningsih, and R. Maulana, “Analisis Perbandingan Algoritma Klasifikasi Data Mining Untuk Dataset Blogger Dengan Rapid Miner,” *J. Khatulistiwa Inform.*, vol. VI, no. 1, pp. 20–28, 2018.