

Klasifikasi Risiko Penyakit pada Ibu Hamil menggunakan Metode *Modified K-Nearest Neighbor* (MKNN)

Yogi Pinanda¹, Wayan Firdaus Mahmudy², Edy Santoso³

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya
Email: ¹yogipinanda333@gmail.com, ²wayanfm@ub.ac.id, ³edy144@ub.ac.id

Abstrak

Ibu hamil perlu menambah pengetahuan untuk mengetahui seberapa besar tingkat risiko terkena penyakit, tertentu dikarenakan rentannya kondisi pada ibu hamil. Klasifikasi tingkat risiko penyakit pada ibu hamil diharapkan dapat membantu pengguna dalam menemukan solusi yang tepat untuk menanggulangnya. Metode klasifikasi yang digunakan untuk mengetahui tingkat risiko penyakit ibu hamil ini menggunakan *Modified K-Nearest Neighbor* (MKNN). Pengklasifikasian tingkat risiko penyakit pada ibu hamil menggunakan metode *Modified K-Nearest Neighbor* (MKNN) dapat mempermudah untuk mendeteksi penyakit berdasarkan faktor-faktor yang ada. Metode *Modified K-Nearest Neighbor* (MKNN) diimplementasikan pada mesin inferensi sistem pakar agar dapat dilakukannya penarikan kesimpulan berdasarkan pengetahuan yang ada. Hasil akurasi sistem yang didapatkan setelah melakukan pengujian yaitu sebesar 85% yang menunjukkan bahwa metode *Modified K-Nearest Neighbor* (MKNN) cocok untuk studi tingkat risiko penyakit pada ibu hamil.

Kata Kunci : Sistem Pakar, Penyakit Pada Ibu Hamil, Klasifikasi, *Modified K-Nearest Neighbor*

Abstract

Pregnant women need to increase their knowledge to find out how big the level of risk of getting a disease, especially because of the vulnerability of pregnant women. Classification of the level of disease risk in pregnant women is expected to assist users in finding the right solution to overcome it. The classification method used to determine the level of disease risk for pregnant women uses Modified K-Nearest Neighbor (MKNN). Classification of disease risk levels in pregnant women using the Modified K-Nearest Neighbor (MKNN) method can make it easier to detect disease based on existing factors. The Modified K-Nearest Neighbor (MKNN) method is implemented on the expert system inference engine so that conclusions can be drawn based on existing knowledge. The results of the accuracy of the system obtained after testing is 85% which indicates that the Modified K-Nearest Neighbor (MKNN) method is suitable for studying the level of disease risk in pregnant women.

Keywords: Expert System, Diseases in Pregnant Women, Classification, Modified K-Nearest Neighbor.

1. PENDAHULUAN

Pada beberapa wilayah di Indonesia masih banyak dijumpai kasus pernikahan dini dibawah umur 20 tahun, dan seringkali dijumpai kasus hamil dibawah umur tersebut. Berdasarkan Kepala BKKBN (Badan Kependudukan dan Keluarga Berencana Nasional) idealnya di umur 20 tahun merupakan usia untuk wanita menikah. Faktor usia muda hanyalah salah satu dari beberapa faktor penyebab terjadinya gangguan kehamilan, salah satu faktor yang penting adalah kurangnya pengetahuan tentang kehamilan bagi calon ibu dan bayinya. Ibu hamil perlu

menambah pengetahuan untuk mengetahui penyakit apa saja yang mungkin bisa diderita olehnya, untuk mendapatkan pengetahuan tersebut dan agar bisa melakukan diagnosa dini penyakit kehamilan, diperlukan suatu sistem yang dapat membantu ibu hamil. Sistem ini nantinya diharapkan bisa membantu ibu hamil, terutama di usia muda yang memiliki resiko gangguan kehamilan cukup tinggi agar dapat menjaga kandungannya dengan baik dan meminimalisir tingkat resiko sang ibu maupun bayi dalam kandungan terkena penyakit.

Dalam berbagai bidang, metode sistem pakar telah banyak digunakan untuk penelitian-

penelitian. Pada bidang kesehatan telah dilakukan penelitian yang mendiagnosa Demam berdarah maupun *Tuberculosis* dengan memakai system pakar (Brla, Y.P, & Takung E.A.S., 2015). Hasil akhir dari investigasi adalah bahwa tes perhitungan berdasarkan sistem pakar memiliki nilai yang sama dengan tes perhitungan manual.

Teknik *K-Nearest Neighbor* (KNN) menggunakan algoritma terawasi untuk mengklasifikasikan hasil *query instance* yang berdasar pada mayoritas dalam kategori KNN. Tujuan dari teknik ini adalah untuk mengklasifikasikan objek baru menggunakan data uji dan atribut (Widodo et al., 2019). Metode ini merupakan contoh pembelajaran berbasis instance dimana data latih disimpan, sehingga proses pengklasifikasian data yang tidak terklasifikasi dapat dengan mudah ditemukan dengan cara membandingkan data tersebut dengan data yang paling mirip pada data uji yang ada. (Larose, 2005)

Penelitian sebelumnya dilakukan oleh Putri (2017) yang bertujuan untuk mengklasifikasikan data baru yang belum diketahui kelas datanya berdasarkan nilai k terdekat. Luaran dari sistem pada penelitian ini adalah temuan diagnosis penyakit pada kucing dengan memanfaatkan data training yang terdiri dari 240 data dengan 14 parameter dan 5 jenis kelainan kulit yang berbeda. Setelah uji dilakukan, maka ditemukan bahwa $k = 1$ memiliki tingkat akurasi tertinggi yaitu 100% serta akurasi terendah ada di 89,668%. Hal ini menandakan bahwa metode Modified KNN dapat diterapkan dalam sebuah penelitian untuk mendiagnosa penyakit kucing (Putri, 2017). Beberapa penelitian lain juga menunjukkan bahwa KNN cukup akurat untuk proses klasifikasi (Putra dkk. 2020, Muflikah dkk., 2020).

Penelitian ini melatarbelakangi peneliti untuk melakukan penelitian tentang klasifikasi risiko penyakit pada ibu hamil dengan dengan memakai *Modified KNN* sebagai metodenya. Penelitian ini dapat menunjukkan tinggi atau rendahnya resiko penyakit yang dialami pada ibu hamil.

2. TINJAUAN PUSTAKA

2.1 Klasifikasi

Data Mining memiliki banyak metode yang dapat digunakan dalam penelitian ini, salah satunya adalah klasifikasi. Tujuan dari metode

ini adalah untuk menebak kelas suatu item yang labelnya tidak diketahui, dan digunakan untuk membangun model yang menjelaskan atau membedakan konsep atau kelas data (Pramudiono, 2006). Dalam hal ini, klasifikasi adalah suatu proses yang dilakukan untuk menebak sebuah objek yang labelnya belum diketahui (Mahmudy dkk., 2021, Anshori dkk., 2019).

2.3 K-Nearest Neighbor

Prosedur klasifikasi dilakukan dengan menghitung kesamaan data latih dengan data uji serta mengkalkulasikan nilai kesamaan k tertinggi. KNN adalah sebuah metode yang paling cocok digunakan untuk melakukan klasifikasi pada suatu objek (Li-juan Zhu, Feng, Z., Qing-qing, P., Xin, Y., & Zheng-tao, Y., 2015).

KNN adalah pendekatan klasifikasi yang menggunakan kedekatan nilai kelas tetangga untuk menentukan klasifikasi. Klasifikasi tersebut didasarkan pada kemiripan atau jarak yang telah disimpan dalam data latih. Algoritma KNN dimulai dengan menghitung nilai K , kemudian menggunakan persamaan Euclidean Distance atau Cosine Similarity (*CosSim*) untuk menghitung nilai kedekatan sekitar antara data uji dan data latih. (D.A. Adeniyi, 2016)

2.4 Modified K-Nearest Neighbor

Algoritma *Modified KNN* (MKNN) merupakan modifikasi dari pendekatan KNN yang mencakup dua proses tambahan yaitu perhitungan bobot dan perhitungan validitas. Algoritma KNN merupakan sebuah algoritma yang terbilang cukup sederhana karena dilakukan dengan hanya mengumpulkan data baru dengan nilai K terdekat (Parvin, H., Alizadeh, H & Bidgoli, B., 2008)

2.5 Nilai Validitas Data Latih

Setiap bagian dari data latih yang digunakan dalam algoritma *K-Nearest Neighbor* yang Dimodifikasi harus divalidasi dan bergantung pada tetangga terdekatnya. Hasil validitas digunakan sebagai informasi tambahan mengenai data setelah dihitung. Persamaan 2.1 menjelaskan rumus untuk menghitung nilai validitas.

Setiap data latih yang akan dihitung pada algoritma *Modified K-Nearest Neighbor* harus divalidasi dan bergantung pada masing-masing

tetangga terdekatnya. Setelah dihitung, hasil validitas digunakan sebagai informasi lebih lanjut tentang data yang dihitung. Rumus untuk menghitung nilai validitas akan dijelaskan pada Persamaan 2.1.

$$Validity(x) = \frac{1}{H} \sum_{i=1}^H S(lbl(x), lbl(Ni(x))) \tag{2.1}$$

Keterangan:

Validity = Validitas antar data latih

H = Jumlah tetangga terdekat

i = Nilai terbaik bernilai 1

lbl(x) = Label kelas *x*

lbl(Ni(x)) = Label kelas titik terdekat dengan *x*

Untuk menentukan kesamaan antara titik *x* dan data ke- *i* dari tetangga terdekat, digunakan fungsi *S*. Persamaan 2.2 akan memberikan rumus untuk menghitung *S*.

$$S(a, b) = \begin{cases} 1 & a = b \\ 0 & a \neq b \end{cases} \tag{2.2}$$

Keterangan:

S = Similarity

a = Kelas a pada data latih

b = Kelas selain a pada data latih

2.6 Perhitungan Jarak Euclidean

Dimungkinkan untuk menentukan jarak antara titik dalam data latih (*x*) dan titik dalam data uji (*y*) menggunakan rumus Euclidean (persamaan 2.3). (*y*).

$$d(x_i, y_i) = \sqrt{\sum_{i=0}^n (x_i - y_i)^2} \tag{2.3}$$

“dengan *d* adalah jarak antara *x* = *x* titik pada data training *x* dan titik data testing *y* yang diklasifikasikan, dimana *x* = *x*₁, *x*₂, ..., *x*_{*i*} dan *y* = *y*₁, *y*₂, ..., *y*_{*i*} dan *i* mewakili nilai atribut dan *n* merupakan atribut dimensi.”

2.7 Weighted Voting

Dalam metode *Modified K-Nearest Neighbor*, dimulai dengan menghitung bobot setiap tetangga menggunakan 1/(*d_e* + 0.5). Setelah itu, pembobotan berdasarkan jarak Euclidean dikalikan dengan validitas masing-masing data pada data latih. Metode ini bekerja dengan baik dengan data yang memiliki nilai validitas lebih tinggi dan paling dekat dengan data. Masalah tersebut dapat diselesaikan dengan mengalikan nilai validitas dengan jarak untuk setiap data dengan jarak dan bobot. Persamaan 2.4 akan menjelaskan rumus untuk menghitung bobot voting.

$$W(i) = Validity(i) \times \frac{1}{d_e + 0.5} \tag{2.4}$$

Keterangan:

W = Bobot antara data uji dengan data latih ke-*i*

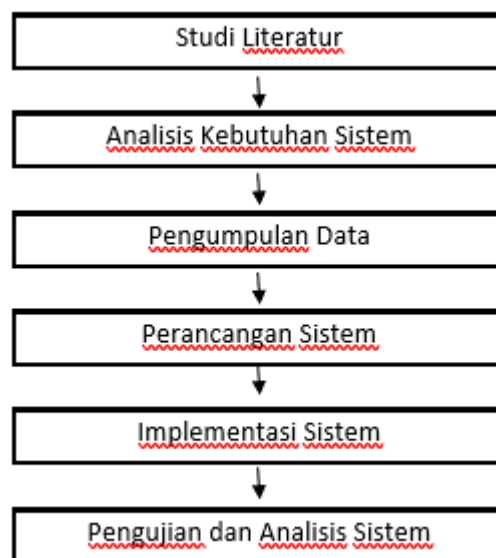
i = Jumlah data latih

Validity = Validitas data latih

d_e = Jarak data latih”

3. METODOLOGI

Pendekatan Modified K-Nearest Neighbor akan digunakan dalam penelitian ini, yang akan dibagi menjadi beberapa tahapan: “tinjauan pustaka, analisis kebutuhan sistem, pengumpulan data, perancangan sistem, implementasi sistem, serta pengujian dan analisis sistem.” Secara umum tahapan sistem digambarkan pada Gambar 1.



Gambar 1 Tahapan Sistem

4. HASIL DAN PEMBAHASAN

4.1 Pengujian Pengaruh Data Latih

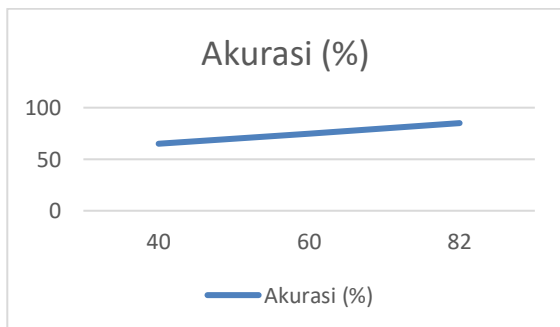
Data latih dari 40, 60, dan 82 dipilih dan digunakan untuk penelitian ini. Tujuan dari pengujian ini adalah untuk melihat besarnya jumlah data latih yang mempengaruhi akurasi. Terdapat sebanyak 20 titik data uji yang digunakan dalam penelitian ini. Dalam pengujian ini, K=3 akan digunakan sebagai nilai K.

Kemudian uji pada data latih akan dilaksanakan yang pada akhirnya, hasil akurasi akan digunakan untuk melihat jumlah data latih akan memengaruhi nilai akurasi yang berbeda-beda. Hasil pengujian pengaruh jumlah data latih yang ditampilkan pada tabel 2 adalah sebagai berikut.

Tabel 1 Hasil Pengaruh Nilai Data Latih

Nilai K	Akurasi (%)
3	85
5	85
7	75

Hasil dari uji coba di atas diperoleh dalam bentuk berbagai jumlah data latih, yang digunakan untuk menentukan akurasi sistem. Akurasi tertinggi diperoleh dengan jumlah training tertinggi yaitu 82 data dengan akurasi 85 persen, sedangkan nilai terendah dihasilkan dengan 40 data training dengan akurasi 65 persen pada pengujian tersebut di atas. Hal ini menunjukkan bahwa semakin banyak data latih maka semakin besar nilai akurasinya. Banyaknya data latih yang digunakan dalam proses klasifikasi akan berdampak pada jumlah data yang digunakan dalam proses klasifikasi. Grafik pengujian yang menggambarkan pengaruh jumlah data latih pada Gambar 2 ditunjukkan di bawah ini.



Gambar 2 Grafik Hasil Pengujian Pengaruh Data Latih

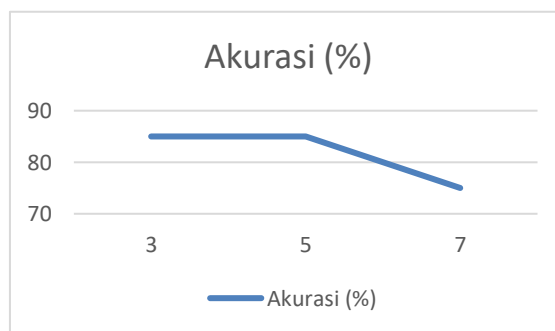
4.2 Pengujian Nilai K

Tujuan pengujian pengaruh nilai K terhadap hasil akurasi adalah untuk melihat seberapa besar pengaruh nilai K terhadap hasil akurasi. Nilai K yang akan digunakan dalam penelitian ini adalah K = 3 sampai K = 7. Pengujian ini dibagi menjadi tiga uji coba, yaitu data uji 25% dari total dataset secara keseluruhan, data uji 50% dari total dataset total, dan terakhir data uji 75% dari total dataset secara keseluruhan. Jumlah total data yang akan digunakan dalam penelitian ini adalah 102 data. Nilai rata-rata akurasi total dari percobaan yang telah dilakukan diambil setelah melakukan setiap percobaan sebanyak 5 kali secara berurutan dengan komposisi data yang berbeda.

Tabel 2 Hasil Pengaruh Nilai K

Nilai K	Akurasi (%)
3	85
5	85
7	75

Setelah percobaan, berbagai hasil akurasi dicapai, dengan akurasi terbaik 85 persen diperoleh pada nilai K = 3 dan K = 5, dan akurasi terendah 75 persen diperoleh pada nilai K = 7. Ini menunjukkan bagaimana nilai K mempengaruhi nilai akurasi; semakin besar nilai K, semakin rendah akurasinya, karena semakin banyak tetangga yang diperhitungkan saat membuat keputusan klasifikasi. Pada Gambar 3, Anda dapat melihat grafik uji nilai K.



Gambar 3 Grafik Hasil Pengujian Nilai K.

5. KESIMPULAN

Berikut adalah kesimpulan yang dapat diambil dari penelitian klasifikasi tingkat risiko penyakit pada ibu hamil menggunakan metode *Modified K-Nearest Neighbor*:

1. Pendekatan Modified K-Nearest Neighbor yang menggunakan 15 parameter dan dua kelas klasifikasi dapat digunakan untuk mengklasifikasikan tingkat risiko penyakit pada ibu hamil.
2. Ketepatan hasil pengujian dalam penelitian ini adalah sebagai berikut:
 - a. Sebuah. Nilai akurasi tertinggi dengan $K = 3$ adalah 85 persen, dan terendah dengan $K = 7$ adalah 75 persen, menurut uji nilai K . Kesimpulan lainnya adalah semakin tinggi nilai K maka semakin rendah tingkat akurasinya karena bertambahnya jumlah tetangga yang akan dijadikan pertimbangan dalam proses klasifikasi.
 - b. Berdasarkan jumlah pengujian terhadap jumlah data latih, nilai akurasi tertinggi adalah 85% dengan 82 data latih dan 20 data uji, kemudian terendah dengan 40 data latih dan 20 data uji dengan akurasi 65%. Kesimpulannya adalah semakin banyak data latih maka semakin besar nilai akurasinya karena semakin banyak data yang harus diperiksa dalam proses klasifikasi.

6 DAFTAR PUSTAKA

- Anshori, M, Mahmudy, WF, Supianto, AA (2019). *Classification Tuberculosis DNA using LDA-SVM*, Journal of Information Technology and Computer Science, vol. 4, no. 3, pp. 233-240.
- Li-juan Zhu, Feng, Z., Qing-qing, P., Xin, Y., & Zheng-tao, Y. (2015). *A classification method of Vietnamese news events based on maximum entropy model*. 2015 34th Chinese Control Conference (CCC).
- Brla, Y.P, & Takung E.A.S. (2015). *Pengembangan Sistem Pakar Diagnosa Penyakit Tuberculosis dan Demam Berdarah Berbasis Web Menggunakan Metode Certainty Factor*.
- Larose, D. T. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining*.
- Mahmudy, WF, Dewi, C, Arifando, R & Rahman, MA (2021), *Combination of Morphology, Wavelet and Convex Hull Features in Classification of Patchouli Varieties with Imbalance Data using Artificial Neural Network*, Journal of Applied Research and Technology, vol. 19, no. 6, pp. 633-643.
- Muflikhah, L, Widodo, Mahmudy, WF & Solimun (2020). *Prediction of Liver Cancer Based on DNA Sequence Using Ensemble Method*, 2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), Yogyakarta, Indonesia, 10-11 December, pp. 37-41.
- Muhammad & Lareno, Bambang. (2015). *Modified Nearest Neighbor Untuk Prediksi Curah Hujan*. Konferensi Nasional Sistem & Informatika.
- Parvin, H., Alizadeh, H & Bidgoli, B. (2008). *MKNN: Modified K-Nearest Neighbor*. San Fransisco. USA: Proceedings of the World Congress on.
- Pramudiono, I. (2006). *Pengantar Data Mining: Menambang Permata Pengetahuan di Gunung Data*. <http://www.ilmukomputer.org/wpcontent/uploads/2006/08/ikodatamining.zip>.
- Putra, FAIA, Utaminingrum, F & Mahmudy, WF (2020). *HOG Feature Extraction and KNN Classification for Detecting Vehicle in The Highway*, Indonesian Journal of Computing and Cybernetics Systems, vol. 14, no. 3, pp. 231-242.
- Putri, M. B. (2017). *Diagnosis Penyakit Kulit Pada Kucing Menggunakan Metode Modified K-Nearest Neighbor*. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN: 2548-964X.
- Ratnasari, Dinda, D., & Sutariyani. (2015). *Sistem Pakar Diagnosa Gangguan Kehamilan Dengan Metode Forward*

Chaining. Jurnal Ilmiah Go Infotech 21(2), 25-31.

Sulistiyano, H., Nurgiyatna, & Madina, A. (2014). *Aplikasi Pendukung Keputusan Pemilihan Konsentrasi Program Studi Bagi Mahasiswa UMS Dengan Menggunakan Analytical Hierarchy Process*. Prosiding Simposium Nasional Teknologi Terapan (SNTT), ISSN:2339-028X.

Syauqy, D. (2018). *Classification of Physical Soil Condition for Plants using Nearest Neighbor Algorithm with Dimensionality Reduction of Color and Moisture Information*. Journal of Information Technology and Computer Science Volume 3, Number 2, 2018, pp. 175-183.

Widodo, AW, Hernando, D & Mahmudy, WF (2019) *Mangrove Forest Classification in Drone Images Using HSV Color Moment and Haralick Features Extraction with K-Nearest Neighbor*, Signal and Image Processing Letters, vol. 1, no. 3, pp. 1-12.