

Implementasi Metode Fuzzy K-Nearest Neighbor dalam Klasifikasi Penyakit Demam Berdarah

Fanny Ramadhani¹, Andy Satria², Indah Purnama Sari³

¹ Fakultas Matematika dan Ilmu Pengetahuan Alam, Ilmu Komputer, Universitas Negeri Medan, Indonesia

² Fakultas Teknik dan Ilmu Komputer, Teknologi Informasi, Universitas Dharmawangsa, Medan, Indonesia

³ Fakultas Ilmu Komputer dan Teknologi Informasi, Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara, Medan, Indonesia



INFORMASI ARTIKEL

Diterima Redaksi: 01 April 2023
Revisi Akhir: 14 April 2023
Diterbitkan Online: 17 Mei 2023

KATA KUNCI

Demam Berdarah Dengue; Klasifikasi; FKNN;
K-Fold Cross Validation; Confusion Matrix

KORESPONDENSI

Phone: +62 82280575321
E-mail: Fannyr@unimed.ac.id

Demam berdarah dengue (DBD) adalah suatu penyakit yang disebabkan oleh infeksi virus dengue. DBD adalah penyakit akut dengan manifestasi klinis perdarahan yang menimbulkan syok yang berujung kematian. DBD disebabkan oleh salah satu dari empat serotipe virus dari genus *Flavivirus*. DBD memiliki gejala yang mirip dengan penyakit lain seperti demam, nyeri sendi, bercak merah di kulit dan lain sebagainya. Maka dari itu diperlukan ketepatan dan akurasi yang baik dalam mengklasifikasikan pasien dbd atau tidak agar tidak terjadi kesalahan dalam memprediksi suatu penyakit. Pada penelitian ini klasifikasi penyakit DBD menggunakan metode Fuzzy K-Nearest Neighbor (FKNN) untuk mencari nilai paling optimum untuk mendapatkan hasil klasifikasi yang paling baik. Dalam penelitian ini dilakukan 8 kali pencarian nilai K, yaitu $k = 3, 5, 7, 9, 11, 13, 15, 20$. Hasil akurasi dari masing-masing K adalah 75,15%, 75,16%, 77,76%, 79,80%, 83,12%, 85,01%, 78,14%, 75,20%. akurasi terbaik didapatkan pada nilai $K = 13$ dengan nilai akurasi mencapai 85,01%

PENDAHULUAN

Demam berdarah dengue (DBD) adalah suatu penyakit yang disebabkan oleh infeksi virus dengue. DBD adalah penyakit akut dengan manifestasi klinis perdarahan yang menimbulkan syok yang berujung kematian. DBD disebabkan oleh salah satu dari empat serotipe virus dari genus *Flavivirus*, famili *Flaviviridae*. [1]. DBD memiliki gejala yang mirip dengan penyakit lain seperti demam, nyeri sendi, bercak merah di kulit dan lain sebagainya. Maka dari itu diperlukan ketepatan dan akurasi yang baik dalam mengklasifikasikan pasien DBD agar tidak terjadi kesalahan dalam memprediksi suatu penyakit.

Klasifikasi bisa digunakan sebagai acuan awal untuk mengidentifikasi apakah orang tersebut masuk pada kelas sakit demam berdarah atau tidak sebelum dilakukan pemeriksaan medis. Klasifikasi adalah pembelajaran secara terlatih (supervised learning). Klasifikasi digunakan untuk memprediksi kelas dari objek yang kelasnya belum diketahui. Metode klasifikasi yang umum digunakan adalah K-Nearest Neighbor, Fuzzy K-Nearest Neighbor (FKNN), Naives Bayes, Neural Network, support vector machines (SVM) dan Decision Tree. [2].

Terdapat salah satu metode artificial intelligence (AI), yang memiliki kinerja yang optimal yakni metode Fuzzy K-Nearest Neighbor (FKNN). Metode Fuzzy K-NN merupakan sebuah metode pengklasifikasian dengan melakukan pencarian nilai keanggotaan pada setiap kelas. Metode FKNN telah banyak digunakan oleh peneliti lainnya dalam mengidentifikasi berbagai penelitian seperti, penelitian yang dilakukan pada tahun 2019 mengenai “Diagnosis Penyakit Cabai Menggunakan Metode Fuzzy K-Nearest Neighbor (FKNN)”, dengan mendapatkan hasil pendiagnosaan penyakit yang menyerang pada tanaman cabai, dengan menggunakan pengujian terhadap nilai K sebanyak 4 kali yakni nilai $K=5$, $K=10$, $K=15$ dan $K=20$, dengan hasil pengujian yang didapatkan yakni akurasi yang optimal sebesar 92% pada nilai $K=5$ [3]. Penelitian lain dalam penggunaan metode FKNN dilakukan pada tahun 2019 mengenai “Implementasi Metode

FKNN (Fuzzy K-Nearest Neighbor) Untuk Diagnosa Penyakit Tanaman Kentang”, dengan mendapatkan hasil berupa pendeteksi penyakit tanaman kentang, dengan dilakukan pengujian sebanyak 4 kali pada nilai $K=5$, $K=10$, $K=15$ dan $K=20$, dengan hasil akurasi yang optimal yakni sebesar 91,4% pada nilai $K=15$ [4]. Penelitian lain juga dilakukan pada tahun 2017 mengenai “Klasifikasi kualitas Tanaman Cabai Menggunakan Metode Fuzzy K-Nearest Neighbor (FKNN)”, dengan menghasilkan hasil akurasi yang optimal sebesar 96,67% dari data sampel sebanyak 100 data dengan pembagian data training dan data testing sebesar 70 untuk data training dan 30 untuk data testing [5]. Dari beberapa penelitian yang menggunakan metode FKNN mendapatkan hasil akurasi yang baik dalam melakukan klasifikasi, sehingga dalam mengidentifikasi penyakit DBD ini menggunakan metode Fuzzy K-Nearest Neighbor (FKNN), untuk mendapatkan hasil klasifikasi yang baik. Dengan dilakukan identifikasi penyakit DBD ini, untuk membantu para tenaga medis dalam mengidentifikasi penyakit DBD dengan cepat, sehingga dapat meminimalisir kematian yang meningkat yang disebabkan oleh penyakit DBD. Penggunaan metode Fuzzy K-Nearest Neighbor (FKNN) pada identifikasi penyakit DBD diharapkan mendapatkan hasil klasifikasi yang lebih baik, tepat dan akurat.

TINJAUAN PUSTAKA

Demam Berdarah Dengue (DBD)

Demam berdarah dengue (DBD) adalah suatu penyakit yang disebabkan oleh infeksi virus dengue. DBD adalah penyakit akut dengan manifestasi klinis perdarahan yang menimbulkan syok yang berujung kematian. DBD disebabkan oleh salah satu dari empat serotipe virus dari genus *Flavivirus*, famili *Flaviviridae*. Setiap serotipe cukup berbeda sehingga tidak ada proteksi-silang dan wabah yang disebabkan beberapa serotipe (hiperendemisitas) dapat terjadi. Virus ini bisa masuk ke dalam tubuh manusia dengan perantara nyamuk *Aedes aegypti* dan *Aedes albopictus*. Terdapat tiga faktor yang memegang peranan pada penularan infeksi virus dengue, yaitu mausia, virus dan vektor perantara. Virus dengue ditularkan kepada manusia melalui nyamuk *Aedes Aegypti*. *Aedes albopictus*, *Aedes polynesiensis* dan beberapa spesies yang lain dapat juga menularkan virus ini, namun merupakan vektor yang kurang berperan. *Aedes* tersebut mengandung virus dengue pada saat menggigit manusia yang sedang mengalami viremia. Adapun gejala dari dbd adalah Demam tinggi mencapai 40 derajat Celsius, Nyeri kepala berat, Nyeri pada sendi, otot, dan tulang, Nyeri pada bagian belakang mata, Nafsu makan menurun, Mual dan muntah, Pembengkakan kelenjar getah bening, Ruam kemerahan sekitar 2–5 hari setelah demam. [1]

Fuzzy K-Nearest Neighbor (FKNN)

Fuzzy K-Nearest Neighbor (FKNN) merupakan sebuah metode yang digunakan dalam pengklasifikasian yang menggabungkan teknik fuzzy dengan K-Nearest Neighbor classifier. Teori himpunan fuzzy dalam menggeneralisasikan teori K-NN klasik yakni mendefinisikan nilai keanggotaan pada sebuah data dalam masing-masing kelas. FKNN yakni sebuah metode yang akan mencari nilai derajat keanggotaan data uji pada setiap kelasnya, yang kemudian akan mengambil nilai derajat keanggotaan terbesar yang dihasilkan. Nilai derajat keanggotaan terbesar akan digunakan sebagai kelas hasil dari klasifikasi.

Proses perhitungan dari metode FKNN akan dimulai dengan menginputkan data latih serta data uji yang akan digunakan dalam proses klasifikasi FKNN, kemudian dilakukan penentuan terhadap jumlah nilai K tetangga terdekat yang akan dilakukan dalam perhitungan. Kemudian dilakukan perhitungan jarak dengan menggunakan fungsi *euclidean distance* pada data uji dan data latih. Pada persamaan yang dapat dilihat pada (2) merupakan rumus perhitungan jarak *euclidean distance* [6].

$$d(X_1, X_2) = \sqrt{\sum_{i=1}^n (X_{i1} - X_{i2})^2} \quad (1)$$

Dalam perhitungan jarak terdekat ada beberapa simbol yakni d_i yang memiliki arti jarak kedekatan (*euclidean distance*), kemudian adanya simbol p yang memiliki arti bahwa jumlah atribut, simbol x_1 yang artinya data latih serta x_2 yang artinya data uji.

Dalam perhitungan FKNN adanya perhitungan pada data latih dalam penentuan kelas data. Pada persamaan yang dapat dilihat pada persamaan (2) merupakan rumus perhitungan penentuan kelas pada data latih [7].

$$u_{ij} = \begin{cases} 0.51 + \frac{n_j}{n} * 0.49, & \text{Jika } j = i \\ \frac{n_j}{n} * 0.49, & \text{Jika } j \neq i \end{cases} \quad (2)$$

Pada penentuan suatu data latih adanya simbol n_j yang mengartikan bahwa jumlah anggota pada kelas j pada data latih n , simbol N mendefinisikan jumlah dari data latih yang digunakan pada penelitian, serta simbol J artinya kelas pada data. Proses perhitungan terakhir pada metode FKNN yakni menentukan keanggotaan setiap kelas. Pada persamaan yang dapat dilihat pada persamaan (3) merupakan perhitungan keanggotaan setiap kelas [8].

$$u_i(x) = \frac{\sum_{j=1}^k u_{ij} \left(\frac{\|x - x_j\|^{-2}}{\|x - x_j\|^{(m-1)}} \right)}{\sum_{j=1}^k \left(\frac{\|x - x_j\|^{-2}}{\|x - x_j\|^{(m-1)}} \right)} \quad (3)$$

Pada rumus perhitungan penentuan keanggotaan setiap kelas terdapat simbol $u_i(x)$ yang mendefinisikan nilai keanggotaan data x dalam kelas ke i , kemudian adanya simbol k yang mendefinisikan jumlah tetangga terdekat yang dimiliki, simbol u_{ij} yang artinya nilai keanggotaan pada kelas i vektor j , simbol $x - x_j$ yang artinya selisih jarak data uji ke data latih, serta simbol m yang artinya bobot pangkat yang dimana bobot pangkat memiliki nilai lebih besar dari pada 1.

Setelah didapatkan hasil perhitungan keanggotaan setiap kelas, maka nilai keanggotaan yang diambil yakni nilai keanggotaan yang terbesar. Hasil dari nilai keanggotaan terbesar merupakan hasil dari klasifikasi penyakit DBD dengan menggunakan metode FKNN.

Teknik Validasi

Cross-validasi (*cross validation*) atau yang sering disebut dengan estimasi rotasi merupakan teknik validasi model untuk menilai keoptimalan hasil analisis, selain itu cross-validasi juga merupakan teknik komposisi dalam penentuan banyaknya data training dan data testing yang akan digunakan. Ada beberapa metode dalam cross-validasi diantaranya yang pertama metode k-fold. Dalam metode k-fold, data disegmentasi secara random ke dalam k partisi yang berukuran sama. Selama proses, salah satu dari partisi dipilih untuk menjadi data testing, sedangkan sisanya digunakan untuk data training. Metode cross-validasi yang kedua yaitu metode holdout. Dalam metode holdout, data awal yang diberi label dipartisi ke dalam dua himpunan secara random yang dinamakan data training dan data testing. Proporsi data yang dicadangkan untuk data training dan data testing tergantung pada analisis misalnya 70%-30% atau 2/3 untuk training dan 1/3 untuk testing, namun menurut [9] serta [10] pada umumnya perbandingan yang digunakan yaitu 2:1 untuk data training berbanding data testing.

Metode Pengujian

Evaluasi dari hasil klasifikasi penyakit DBD dilakukan dengan *confusion matrix*. Metode ini merepresentasikan hasil klasifikasi menggunakan matriks yang dapat dilihat pada Tabel 1.

Tabel 1. Confusion Matrix		
Predicted Value	Actual Value	
	TP	FP
	FN	TN

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} * 100\% \quad (2)$$

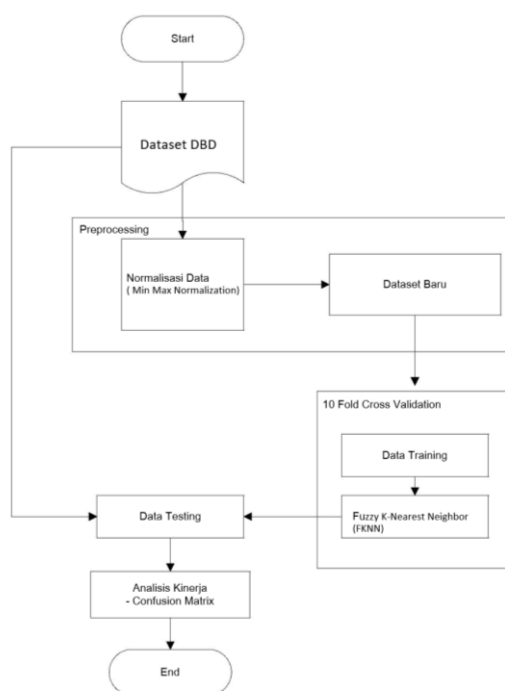
METODOLOGI

Penelitian ini dilakukan dengan beberapa tahapan yaitu mulai dari proses studi literatur, pengumpulan data, klasifikasi menggunakan metode fuzzy K-nearest Neighbor (FKNN) dan pengujian. Studi literatur digunakan untuk menguatkan permasalahan yang diangkat dari beberapa sumber yang berkaitan dengan masalah yang di angkat. Kemudian dilanjutkan dengan mengumpulkan data. Data yang digunakan dalam penelitian ini menggunakan data sekunder yang di dapat dari UCI machine learning. Data terdiri dari 300 data. Kemudian pada proses preprocessing data dilakukan dengan proses normalisasi data untuk standarisasi data agar pada saat pengujian tidak ada yang mendominasi data. Metode normalisasi

yang digunakan adalah metode Min-Max Normalization agar skala nilai hanya memiliki rentang 0 sampai 1 pada semua data. Rumus perhitungan min max normalization dapat dilihat pada persamaan 5.

$$V' = \frac{v - \min A}{\max A - \min A} \quad (3)$$

Setelah data melewati tahapan preprocessing data pada proses normalisasi, proses selanjutnya adalah pembagian data dengan menggunakan k-fold cross validation pada dataset DBD. Penggunaan k-fold cross validation dalam pembagian data dikarenakan hasil pengujian akan lebih optimal dan baik, sebab semua dataset yang digunakan pada penelitian akan digunakan sebagai data training dan data testing pada saat pengujian dilakukan [11][12]. Setelah dilakukan pembagian data, selanjutnya dilakukan klasifikasi dengan menggunakan metode Fuzzy K-Nearest Neighbor (FKNN). Setelah klasifikasi menggunakan metode Fuzzy K-Nearest Neighbor (FKNN), dilakukan pengujian menggunakan confusion matrix dalam melakukan perhitungan tingkat akurasi klasifikasi penyakit DBD berdasarkan data yang telah digunakan. Pada gambar yang dapat dilihat pada Gambar 1 merupakan flowchart klasifikasi FKNN pada data DBD.



Gambar 1. Flowchart Klasifikasi FKNN Pada Data DBD

HASIL DAN PEMBAHASAN

Penelitian ini melakukan klasifikasi pada penyakit DBD dengan metode Fuzzy K-Nearest Neighbor (FKNN). Data yang digunakan dalam penelitian ini menggunakan data sekunder yang di dapat dari UCI machine learning. Data terdiri dari 300 data, 13 atribut dan 2 class. Pembagian data dalam penelitian klasifikasi penyakit DBD dengan metode Fuzzy K-Nearest Neighbor (FKNN) menggunakan k-fold cross validation, dimana dataset dibagi menjadi data training dan data testing Pada perhitungan FKNN, pengolahan data dibagi menjadi 70 % data training dan 30% data testing. Dimana terdapat 240 data training dan 60 testing. Pembagian data didapat dari proses 10- *Fold Cross Validation*. Pada tahap pertama, mencari nilai K terbaik. Pengujian terhadap nilai K yang digunakan pada penelitian ini yakni pengujian pada K=3, K=5, K=7, K=9, K=11, K=13 dan K=20. Hasil nilai akurasi dengan K berbeda dapat dilihat pada tabel 2 dibawah ini.

Tabel 2. Akurasi dengan K berbeda

Akurasi	
K= 3	75,15%
K= 5	75.16%
K=7	77.76%
K=9	79.80%

K=11	83.12%
K=13	85.01%
K=15	78.14%
K=20	75.20%

Perhitungan hasil akurasi yang didapatkan pada penelitian menggunakan perhitungan metode confusion matrix. Dari beberapa hasil akurasi yang didapatkan penggunaan metode Fuzzy KNearest Neighbor (FKNN) mendapatkan hasil yang baik dalam klasifikasi penyakit DBD. Pada data diatas dapat dilihat bahwa akurasi terbaik didapatkan pada nilai K= 13 dengan nilai akurasi mencapai 85.01%.

KESIMPULAN DAN SARAN

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa penggunaan metode Fuzzy K-Nearest Neighbor (FKNN) pada klasifikasi penyakit DBD, dengan menggunakan pembagian dataset 70% data training serta 30% data testing pada 300 jumlah dataset mendapatkan hasil akurasi yang optimal pada K=13 dengan hasil akurasi sebesar 85.01% pada pengujian.

DAFTAR PUSTAKA

- [1] Sukohar. (2014). Demam Berdarah Dengue (DBD). *Fever is to Combine The "3M Plus", Which is closing, Draining and Hoarding. Treatment of Patients with dengue fever is symptomatic and supportive*, pp. 1-15.
- [2] Richard O.Duda, E. D. (2006). *Pattern Classification*. Canada: John Wiley & Sons, Inc.
- [3] W. G. Akbari, N. Hidayat, and N. Santoso, "Diagnosis Penyakit Cabai Menggunakan Metode Fuzzy K-Nearest Neighbor (FKNN)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 3, no. 1, pp. 1070–1074, 2019.
- [4] D. M. F. Shanti, N. Hidayat, and R. C. Wihandika, "Implementasi Metode F-KNN (Fuzzy K-Nearest Neighbor) Untuk Diagnosis Penyakit Anjing," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 12, pp. 7401–7407, 2018.
- [5] A. A. Indra Wiratmaka, I. F. Rozi, and R. A. Asmara, "Klasifikasi Kualitas Tanaman Cabai Menggunakan Metode Fuzzy K-Nearest Neighbor (Fknn)," *J. Inform. Polinema*, vol. 3, no. 3, p. 1, 2017, doi: 10.33795/jip.v3i3.25.
- [6] Dylan Trotsek, "APLIKASI KLASIFIKASI JENIS BAJA BERDASARKAN KOMPOSISI KIMIA DENGAN MENGGUNAKAN METODE FK-NNC," *J. Chem. Inf. Model.*, 2017.
- [7] C. Zheng, M. Shereduwan, M. Kasihmuddin, M. A. Mansor, and J. Chen, "Intelligent Multi-Strategy Hybrid Fuzzy K-Nearest Neighbor Using Improved Hybrid Sine Cosine Algorithm," pp. 1–23, 2022.
- [8] D. W. I. Ispriyanti, A. Prahutama, R. Dian, and I. K. A. Wati, "Analysis classification of households who received 'raskin' in Semarang City using Fuzzy K-Nearest Neighbor (FKNN) and Support Vector Machine (SVM)," *J. Math. Comput. Sci.*, pp. 1–13, 2022, doi: 10.28919/jmcs/7478.
- [9] Han, J & Kamber, M. (2006). *Data Mining Concepts And Techniques, Second Edition*. California: Morgan Kaufman.
- [10] Witten, I.H & Frank, E. (2005). *Data Mining Practical Machine Learning Tools And Teachniques. Second edition*. California: Morgan Kaufman.
- [11] A. Satria, O. S. Sitompul and H. Mawengkang, "5-Fold Cross Validation on Supporting K-Nearest Neighbour Accuration of Making Consimilar Symptoms Disease Classification," in *IEEE*, Padang, 2021.
- [12] F. Ramadhani, Al-Khowarizmi and I.P. Sari, "Improving the Performance of Naïve Bayes Algorithm by Reducing the Attributes of Dataset Using Gain ratio and Adaboost," in *EEE*, Padang, 2021.
- [13] I. P. Sari, I. H. Batubara and Al-Khowarizmi, "Sensitivity of Obtaining Errors in The Combination of Fuzzy and Neural Networks for Conducting Student Assessment on E-Learning," *International Journal of Economic, Technology and Social Sciences (Injects)*, vol. 2, no. 1, pp. 331 - 338, 2021.