

Literatur Review: Klasifikasi Penyakit Parasit dengan Algoritma Decision Tree dan K-Nearest Neighbors (KNN)

Muhammad Mabdail Hidayat^{1*}, Kakan Kandarsyah¹, Randy Rizkiani¹,
Fendi Indra Pradana¹

¹Fakultas Ilmu Komputer, Teknik Informatika, Universitas Pamulang, Jl. Raya Puspiptek No. 46,
Kel. Buaran, Kec. Serpong, Kota Tangerang Selatan. Banten 15310, Indonesia

Email: ^{1*}muhammadmabdail@gmail.com, ²Kandarsyah89@gmail.com,
³randyrizkiani5@gmail.com, ⁴indrapradan728@gmail.com,

(* : coresponding author)

Abstrak– Kajian literatur ini membahas klasifikasi penyakit parasit menggunakan algoritma Decision Tree dan K-Nearest Neighbors (KNN). Penyakit parasit, yang banyak ditemukan di daerah tropis, membutuhkan diagnosa yang tepat untuk mencegah penyebaran dan meningkatkan efektivitas pengobatan. Dalam beberapa dekade terakhir, algoritma Decision Tree dan KNN telah digunakan secara luas dalam klasifikasi data medis, khususnya untuk diagnosa penyakit. Studi ini bertujuan untuk mengevaluasi keefektifan kedua algoritma ini dalam klasifikasi penyakit parasit berdasarkan kajian literatur terkini. Metode kajian literatur dilakukan dengan mengumpulkan dan menganalisis lima artikel terkait dalam lima tahun terakhir. Hasil menunjukkan bahwa kedua algoritma memiliki kelebihan dan kelemahan masing-masing; KNN unggul dalam akurasi pada dataset besar sementara Decision Tree memberikan interpretasi hasil yang lebih mudah. Tantangan utama dalam penggunaan kedua algoritma ini melibatkan pemilihan parameter dan sensitivitas data. Rekomendasi lebih lanjut dalam studi ini mencakup penggunaan teknik ensemble untuk menggabungkan kelebihan kedua algoritma.

Kata Kunci: Klasifikasi Penyakit, *Decision Tree*, *K-Nearest Neighbors*, Penyakit Parasit, *Machine Learning*

Abstract– This literature review discusses the classification of parasitic diseases using Decision Tree and K-Nearest Neighbors (KNN) algorithms. Parasitic diseases, which are commonly found in tropical areas, require accurate diagnosis to prevent their spread and improve the effectiveness of treatment. In recent decades, Decision Tree and KNN algorithms have been widely used in medical data classification, especially for disease diagnosis. This study aims to evaluate the effectiveness of these two algorithms in parasitic disease classification based on a recent literature review. The literature review method was carried out by collecting and analyzing five related articles in the last five years. The results show that both algorithms have their own advantages and disadvantages; KNN excels in accuracy on large datasets while Decision Tree provides easier interpretation of results. The main challenges in using these two algorithms involve parameter selection and data sensitivity. Further recommendations in this study include the use of ensemble techniques to combine the advantages of both algorithms.

Keywords: Disease Classification, *Decision Tree*, *K-Nearest Neighbors*, Parasitic Disease, *Machine Learning*

1. PENDAHULUAN

Penyakit parasit masih menjadi ancaman kesehatan yang signifikan, terutama di negara berkembang dan wilayah tropis. Penyakit ini tidak hanya mengakibatkan penurunan kualitas hidup tetapi juga dapat menyebabkan kematian jika tidak diobati dengan cepat. Diagnosis penyakit parasit secara manual membutuhkan waktu dan sumber daya yang signifikan, yang dapat menyebabkan keterlambatan dalam penanganan. Seiring dengan perkembangan teknologi, metode machine learning mulai diterapkan dalam dunia medis, salah satunya adalah untuk mendeteksi dan mengklasifikasikan penyakit parasit secara otomatis.

Algoritma Decision Tree dan K-Nearest Neighbors (KNN) adalah dua metode yang sering digunakan dalam klasifikasi data medis. Decision Tree dikenal karena kemampuannya untuk menghasilkan model yang mudah diinterpretasi, yang sangat penting dalam konteks medis. Algoritma ini membagi data ke dalam kategori yang jelas berdasarkan aturan logika sederhana, sehingga hasil diagnosa dapat lebih mudah dipahami oleh profesional medis. Di sisi lain, KNN adalah algoritma berbasis instance yang menghitung jarak antara titik data untuk menentukan kelasnya. KNN dikenal dengan kemampuannya untuk memberikan hasil yang akurat pada dataset besar, meskipun membutuhkan waktu komputasi yang lebih tinggi.

Dalam penelitian ini, dilakukan kajian literatur terhadap studi-studi yang memanfaatkan Decision Tree dan KNN untuk klasifikasi penyakit parasit. Kajian ini bertujuan untuk mengevaluasi kelebihan dan kelemahan dari masing-masing algoritma, serta untuk memberikan rekomendasi terhadap penerapan yang lebih efektif dalam klasifikasi penyakit parasit.

2. METODOLOGI PENELITIAN

2.1 Systematic Literatur Review (SLR)

Tujuan dari Tinjauan Pustaka Sistematis (SLR) adalah untuk mengumpulkan, menilai, dan mensintesis literatur secara metodis yang berkaitan dengan subjek penelitian. Dengan menggunakan istilah "Penyakit Parasit", "Pohon Keputusan dalam Klasifikasi", dan "K-Nearest Neighbours", kami mengumpulkan publikasi terkait dari database seperti ScienceDirect, Google Scholar, dan IEEE Xplore untuk penelitian kami.

2.2 Kriteria Inklusi dan Eksklusi

- Kriteria Inklusi: Artikel yang diterbitkan dalam 5 tahun terakhir, menggunakan algoritma Decision Tree atau KNN dalam klasifikasi penyakit parasit, dan memiliki data evaluasi yang relevan (seperti akurasi, presisi, dan recall).
- Kriteria Eksklusi: Artikel yang tidak secara spesifik membahas penggunaan Decision Tree atau KNN dalam klasifikasi penyakit parasit atau yang tidak menyediakan hasil evaluasi model.

2.3 Pengumpulan Data

Informasi dikumpulkan dengan melakukan pencarian kata kunci dan kemudian mengurutkan hasilnya berdasarkan kriteria inklusi dan eksklusi. Untuk mengidentifikasi tren dan pola penggunaan algoritme, kami menganalisis dan mensintesis artikel yang memenuhi persyaratan.

2.4 Pertanyaan Penelitian

Kajian ini difokuskan pada tiga pertanyaan penelitian utama:

1. Bagaimana performa algoritma Decision Tree dibandingkan dengan KNN dalam klasifikasi penyakit parasit?
2. Apa saja kelebihan dan kekurangan dari kedua algoritma tersebut dalam konteks klasifikasi penyakit parasit?
3. Apa metode yang dapat digunakan untuk meningkatkan akurasi dan efisiensi klasifikasi menggunakan Decision Tree dan KNN?

3. ANALISA DAN PEMBAHASAN

Kajian literatur ini mengumpulkan data dari lima artikel yang relevan, dan tabel di bawah ini merangkum metode, tujuan, hasil, dan tahun publikasi masing-masing artikel.

Tabel 1. Hasil Temuan Artikel Relevan

| No | Peneliti dan Tahun | Metode Yang di bahas | Tujuan penelitiannya | Hasil yang didapat |
|----|---|----------------------------------|--|---|
| 1 | Fikry ubaidillah, M.Rizki Ramadhan, Perani Rosyani (2023) | Decision tree dan algoritma C4.5 | Guna mendapatkan faktor utama yang paling berpengaruh terhadap perubahan cuaca | Studi tersebut menemukan bahwa pohon keputusan dengan atribut kelembaban sebagai simpul akarnya mencapai tingkat akurasi sebesar 81,94%, sebagaimana dikonfirmasi oleh alat Rapid Miner 9.10, menggunakan 2.400 data prakiraan cuaca yang diambil dari situs Accuweather. Data dibagi menjadi dua |

| | | | | |
|---|---|--|---|--|
| | | | | set: data pelatihan yang terdiri dari 1.680 record dan data pengujian yang terdiri dari 720 record. |
| 2 | Agustinus Marcello Soebiantoro , Perani Rosyani (2023) | Pohon Klasifikasi, Naive Bayes, dan K-Nearest Neighbors | Memprediksi hujan dengan parameter validasi berupa Skor Brier, Matriks Kebingungan, dan kurva ROC. | Setelah serangkaian pengujian termasuk pemrosesan data, pemilihan model, dan akurasi, Algoritma Naive Bayes muncul sebagai pemenang dengan skor 77,1% dalam area klasifikasi yang cukup baik, yang menunjukkan bahwa Algoritma Naive Bayes mungkin berguna dalam pengoperasian. Faktor meteorologi terpenting yang berkontribusi terhadap perkembangan curah hujan adalah kelembaban relatif (RHavg), suhu terendah (Tmin), suhu maksimum (Tmax), suhu rata-rata (Tavg), dan arah angin (ddd). |
| 3 | Rani Puspita , Agus Widodo (2020) | KNN, Decision Tree, dan Naïve Bayes | Tujuan dari penelitian yang dilakukan ini adalah untuk mengetahui tingkat, akurasi, dari, tiga metode, berbeda diantaranya KNN, Decision Tree dan Naïve, Bayes, dengan menggunakan tools RapidMiner | Hasil penelitian menunjukkan bahwa teknik analisis sentimen KNN memberikan layanan BPJS dengan tingkat akurasi 95,58%, tingkat prediksi positif 45,00%, tingkat prediksi positif 0,00%, dan tingkat prediksi netral 96,83%. Teknik pohon keputusan mencapai tingkat akurasi sebesar 96,13% dengan presisi kelas sebesar 55,00% untuk pred. negatif, 0,00% untuk pred. positif, dan 97,28% untuk pred. netral. Terakhir, teknik Naïve Bayes memperoleh akurasi sebesar 89,14% dengan presisi kelas sebesar 16,67% untuk prediksi negatif, 1,64% untuk prediksi positif, dan 98,40% untuk prediksi netral. |
| 4 | Fitri Anggraeni, Niko Kristiawan, Rikha Lutfiati, Yudha Dirgantara, Perani Rosyani (2023) | Classification Tree | Memprediksi hujan dengan parameter uji Confusion Matrix | Dengan kategori klasifikasi layak, temuan menunjukkan bahwa metode Pohon Klasifikasi mempunyai akurasi sebesar 74,7%. Dari total 1.095 dataset yang diperiksa, terdapat 818 dataset yang memiliki prediksi valid. Hujan terbentuk ketika kondisi meteorologi tertentu terpenuhi, yang terpenting adalah kelembaban rata-rata (RHavg), suhu terendah (Tmin), suhu maksimum (Tmax), suhu rata-rata (Tavg), dan arah angin (ddd). |
| 5 | Ahmad Rizal , Firdha Rofika Bryliana , Krisna Nur Aedi Aripin , Sabita | Association Rule, C4.5, Classification dan Random Forest | Prakiraan cuaca | Hasil penelitian menunjukkan Association Rule mempunyai tingkat akurasi 60.9%, sedangkan C4.5 mempunyai tingkat akurasi 68.5%. Dengan demikian model prediksi yang dipilih adalah model prediksi C4.5. Komponen cuaca yang dominan memungkinkan terjadinya hujan adalah |

| | | | | |
|--|--|--|--|--|
| | Adelia Wardani, Perani Rosyani. (2023) | | | suhu udara, suhu titik embun, dan tutupan awan |
|--|--|--|--|--|

4. KESIMPULAN

Berdasarkan kajian literatur ini, kedua algoritma, Decision Tree dan KNN, memiliki kelebihan dan kelemahan masing-masing dalam klasifikasi penyakit parasit. Decision Tree sangat berguna untuk menghasilkan model yang mudah dipahami, sementara KNN cenderung lebih akurat pada dataset besar. Namun, terdapat beberapa tantangan dalam penerapan kedua algoritma ini, seperti risiko overfitting pada Decision Tree dan tingginya kebutuhan komputasi pada KNN.

Untuk meningkatkan performa, direkomendasikan untuk menggunakan teknik ensemble yang menggabungkan Decision Tree dan KNN, atau menerapkan metode lain seperti Random Forest atau teknik bagging. Selain itu, penelitian lebih lanjut juga dapat dilakukan untuk mengoptimalkan parameter KNN dan mengurangi risiko overfitting pada Decision Tree.

REFERENCES

- Ramadhan, M. R., & Rosyani, P. (2023). IMPLEMENTASI DATA MINING PADA DATASET PRAKIRAAN CUACA MENGGUNAKAN ALGORITMA C4. 5. *ALKHAWARIZMI: Jurnal Matematika, Algoritma dan Sains*, 1(1), 131-140.
- Soebiantoro, A. M., & Rosyani, P. (2023). Aplikasi Data Mining Pada Prediksi Cuaca Menggunakan Klasifikasi Naïve Bayes. *NEWTON: Jurnal Matematika, Fisika, Algoritma dan Sains*, 1(1), 42-45.
- Puspita, R., & Widodo, A. (2021). Perbandingan Metode KNN, Decision Tree, dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS. *Jurnal Informatika Universitas Pamulang*, 5(4), 646-654.
- Anggraeni, F., Kristiawan, N., Lutfiati, R., Dirgantara, Y., & Rosyani, P. (2023). Prediksi Cuaca Yang Akan Datang Menggunakan Metode Data Mining. *NEWTON: Jurnal Matematika, Fisika, Algoritma dan Sains*, 1(1), 73-83.
- Rizal, A., Bryliana, F. R., Aripin, K. N. A., Wardani, S. A., & Rosyani, P. (2023). PEMANFAATAN DATA MINING UNTUK PRAKIRAAN CUACA. *NEWTON: Jurnal Matematika, Fisika, Algoritma dan Sains*, 1(1), 34-41.
- Nurkholifah, M., & Umar, Y. (2023). Analisa Performa Algoritma Machine Learning Dalam Prediksi Penyakit Liver. *Jurnal Indonesia: Manajemen Informatika dan Komunikasi*, 4(1), 164-172.
- Herisnan, D. N. (2024). KOMPARASI ALGORITMA DECISION TREE, SVM, NAIVE BAYES DALAM PREDIKSI PENYAKIT LIVER. *JSR: Jaringan Sistem Informasi Robotik*, 8(1), 104-107.
- Fidyaningsih, S., Agus, F., & Maharani, S. (2016, September). Sistem Pakar Diagnosa Penyakit Kucing Menggunakan Metode Case-Based Reasoning. In *Prosiding Seminar Ilmu Komputer dan Teknologi Informasi* (Vol. 1, No. 1).
- Mardiani, E., Rahmansyah, N., Ningsih, S., Lantana, D. A., Wirawan, A. S. P., Wijaya, S. A., & Putri, D. N. (2023). Komparasi Metode Knn, Naive Bayes, Decision Tree, Ensemble, Linear Regression Terhadap Analisis Performa Pelajar Sma. *Innovative: Journal Of Social Science Research*, 3(2), 13880-13892.
- Marutho, D. (2019). Perbandingan Metode Naive Bayes, KNN, Decision Tree Pada Laporan Water Level Jakarta. *Jurnal Ilmiah Infokam*, 15(2).