**MODEL KOMPUTASIONAL UNTUK MENDIAGNOSIS DIABETES: TINJAUAN SISTEMATIS**

**Alaikassalam, Ivanka Ramadhan, Bilqizah Rahma Ilayya Syahdewi, Adhitiya Aprillio Pratama, Sirojul Izzeh**

**Prodi S1 Teknik Informatika Fakultas Teknik**

**Universitas Trunojoyo Madura**

**Email :** [230411100133@student.trunojoyo.ac.id](mailto:230411100133@student.trunojoyo.ac.id) ; [230411100202@student.trunojoyo.ac.id](mailto:230411100202@student.trunojoyo.ac.id) ; [230411100030@student.trunojoyo.ac.id](mailto:230411100030@student.trunojoyo.ac.id) ; [230411100201@student.trunojoyo.ac.id](mailto:230411100201@student.trunojoyo.ac.id) ; [230411100131@student.trunojoyo.ac.id](mailto:230411100131@student.trunojoyo.ac.id)

**Abstrak**

**Pendahuluan:** Diabetes Melitus (DM) merupakan penyakit kronis yang semakin meningkat prevalensinya di seluruh dunia, termasuk Indonesia, di mana diperkirakan jumlah pengidapnya akan mencapai 16,2 juta jiwa pada tahun 2040. Penyakit ini disebabkan oleh kadar glukosa darah yang tinggi dan dapat memicu berbagai komplikasi serius jika tidak ditangani dengan baik.

**Tujuan:** Penelitian ini bertujuan untuk melakukan tinjauan sistematis terhadap model komputasional yang digunakan untuk mendiagnosis diabetes, dengan fokus pada metode Decision Tree, K-Nearest Neighbors, Naive Bayes, dan Logistic Regression.

**Metode:** Penelitian dilakukan dengan mengumpulkan artikel yang telah ditinjau sejawat dari berbagai sumber, menggunakan kata kunci terkait diagnosis diabetes dan metode komputasional. Data yang digunakan dalam analisis diambil dari dataset publik yang relevan.

**Hasil:** Dari Sekitar 157 hasil paper yang ditemukan, 20 artikel memenuhi kriteria inklusi. Hasil menunjukkan bahwa Logistic Regression dan Decision Tree memiliki akurasi yang lebih tinggi dibandingkan dengan K-Nearest Neighbors dan Naive Bayes, dengan akurasi rata-rata masing-masing 89% dan 87%.

**Kesimpulan:** Meskipun semua metode memiliki kelebihan dan kelemahan, Logistic Regression dan Decision Tree menunjukkan potensi yang lebih baik dalam diagnosis diabetes. Penelitian lebih lanjut diperlukan untuk meningkatkan akurasi dan kecepatan diagnosis.

**Kata Kunci:** Model Komputasi, Diagnosis Diabetes, Decision Tree, K-Nearest Neighbors, Naive Bayes, dan Logistic Regression.

## 1. PENDAHULUAN

Diabetes merupakan salah satu masalah kesehatan global yang terus meningkat. Menurut data WHO, jumlah penderita diabetes diperkirakan mencapai 422 juta pada tahun 2014, dan angka ini terus meningkat. Diabetes dapat menyebabkan berbagai komplikasi serius, termasuk penyakit jantung, kerusakan saraf, dan kebutaan. Oleh karena itu, diagnosis yang tepat dan cepat sangat penting untuk pengelolaan penyakit ini. Berbagai metode komputasional telah dikembangkan untuk membantu dalam diagnosis diabetes, termasuk Decision Tree, K-Nearest Neighbors, Naive Bayes, dan Logistic Regression. Penelitian ini bertujuan untuk menganalisis dan membandingkan efektivitas metode-metode tersebut dalam mendiagnosis diabetes.

Beberapa penelitian terdahulu telah membahas pengembangan sistem pakar untuk diagnosis penyakit diabetes. Widodo et al. (2021) merancang sistem pakar diagnosis penyakit diabetes berbasis web menggunakan algoritma Naive Bayes dengan tingkat akurasi yang cukup tinggi. Husada (2016) mengembangkan aplikasi clinical pathway penyakit kandungan pada ibu hamil menggunakan algoritma Naive Bayes dengan tingkat akurasi mencapai 83,3%. Andryanto et al. (2018) mengimplementasikan metode Naive Bayes untuk menentukan penyakit diabetes mellitus dengan tingkat akurasi terbaik mencapai 100% menggunakan 140 data training.

Diabetes adalah penyakit yang kompleks dan rumit. Tingkat diagnosa diabetes memberikan kontribusi yang signifikan terhadap komorbiditas dan tingkat komplikasi diabetes (CDA, 2008). Berdasarkan data histori penderita penyakit diabetes dapat dibuat rekomendasi prediksi penyakit diabetes yang membantu tenaga kesehatan yaitu menggunakan klasifikasi data dengan decision tree.

Penelitian menggunakan regresi logistik yang dilakukan oleh Gunawan et al dalam menghasilkan prediksi akurasi penyakit diabetes melitus sebesar 72,22% pada regresi logistik tanpa grid search, sedangkan prediksi regresi logistik dengan grid search menghasilkan akurasi sebesar 83,33%. Penelitian lain yang dilakukan oleh Marna et al dalam menghasilkan besarnya peluang faktor seorang mahasiswa yang memiliki ayah yang lulusan SMP dan ibu yang pendidikannya lulusan SMA (eksternal) memiliki waktu belajar kurang lebih 7 jam dan mahasiswa yang bersikap baik (internal) memperoleh IPK dibawah 3 sebesar 0,47 atau 47% sedangkan peluang memperoleh IPK lebih dari 3 adalah 0,53 atau 53%.

K-Nearest Neighbor (KNN) merupakan salah satu algoritma dengan teknik lazy learning yang menggunakan metode klasifikasi terhadap objek berdasarkan data yang jaraknya paling dekat dengan objek tersebut. KNN juga termasuk dalam kelompok instance-based learning. Pada algoritma KNN, data berdimensi q, jarak dari data tersebut ke data yang lain dapat dihitung. nilai dari jarak inilah yang digunakan sebagai nilai kedekatan antara data uji dengan data latih. Terdapat berbagai macam metode yang dapat diterapkan pada algoritma KNN dalam menghitung jarak kedekatan objek dengan data terdekat, namun yang paling umum digunakan adalah metode Euclidean Distance, bahkan digunakan sebagai default pada library SKLearn pada bahasa pemrograman Python saat menggunakan algoritma KNN. Metode penghitungan jarak lainnya yang umum digunakan pada algoritma KNN yaitu Minkowski Distance dan Manhattan Distance.

## 2. BAHAN DAN METODE

Systematic Literature Review (SLR) digunakan sebagai metodologi dalam penelitian ini. Proses-proses yang terkandung dalam metodologi adalah sebagai berikut: desain penelitian, strategi pencarian dan sumber informasi; seleksi studi dan proses pengumpulan data; serta penilaian kualitas dan sintesis data.

### 2.1 Desain Penelitian, Strategi Pencarian dan Sumber Informasi

Secara sistematis, kami meninjau studi berbasis komputasional untuk diagnosis diabetes yang memenuhi kriteria inklusi dan eksklusi yang telah ditentukan sebelumnya. Strategi pencarian kami disusun sebagai berikut: (a) Membangun istilah pencarian dengan mengidentifikasi kata kunci utama; (b) Menentukan sinonim atau kata alternatif untuk kata kunci utama; (c) Menetapkan kriteria eksklusi untuk melakukan penyeleksian selama pencarian; dan (d) Menerapkan operator Boolean untuk menyusun istilah pencarian yang diperlukan.

Hasil untuk (a): Komputasional, Diagnosis, Diabetes.

Hasil untuk (b): Komputasional, Matematika, Decision Tree, K-Nearest Neighbor, Naive Bayes, Logistic Regression.

Hasil dari (c): Non-komputasional, Obat, Hewan, Pencitraan, Radiografi, Vaksin, Tomografi, Finansial.

Hasil untuk (d): Diagnosis intitle: Diabetes Komputasional | Decision Tree | "K-Nearest Neighbor" | "Naive Bayes" | "Logistic Regression" -Obat -Hewan -Pencitraan -Radiografi -Vaksin -"Tomografi" -Finansial.

Dalam penelitian ini, seleksi dilakukan dari artikel yang telah ditinjau sejawat pada database Google Scholar, IEEE Xplore, dan PubMed dengan fokus pada model komputasional untuk diagnosis diabetes menggunakan istilah pencarian yang disusun sebagai "Hasil untuk (d)" dan ini menjadi istilah pencarian final untuk penelitian ini. Sumber yang diperiksa meliputi: prosiding konferensi, artikel jurnal, bab buku, dan buku.

### 2.1.1 Seleksi Studi dan Proses Pengumpulan Data

Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) diadopsi untuk merinci proses seleksi studi yang menetapkan studi yang disertakan dan yang dikecualikan untuk penelitian ini. Studi dikategorikan berdasarkan jenis metode komputasional yang digunakan. Risiko bias dinilai pada tingkat studi. Gambar 1 menyajikan proses seleksi studi dalam diagram alir PRISMA.

[Gambar 1: Diagram alir PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) dari studi yang disertakan]

Studi tentang diagnosis Diabetes menggunakan metode komputasional dipertimbangkan. Studi dibatasi pada diagnosis diabetes pada manusia. Dengan demikian istilah pencarian meliputi: Diabetes, Diagnosis, Komputasional, Decision Tree, K-Nearest Neighbor, Naive Bayes, dan Logistic Regression.

### 2.1.2 Kriteria Eksklusi

Artikel yang diterbitkan diidentifikasi dengan mencari database Google Scholar, IEEE Xplore, dan PubMed hingga tahun 2023. Publikasi non-Bahasa Indonesia tidak dikecualikan. Studi dikecualikan jika didasarkan pada hal-hal berikut: a. Data empiris saja b. Analisis klinis saja c. Eksperimen/prosedur pembelajaran d. Desain studi yang tidak lengkap/tidak sesuai e. Studi pada hewan/non-manusia

### 2.1.3 Pemilihan Sumber Utama

Pemilihan awal sumber utama diambil dari hasil studi yang diidentifikasi berdasarkan istilah pencarian dan memiliki elemen kata kunci yang terkait dengan penelitian ini. Pemilihan juga dilakukan berdasarkan penilaian kualitas publikasi. Penilaian kualitas publikasi dilakukan berdasarkan hal-hal berikut dengan tujuan untuk meminimalkan bias: Apakah penyakit yang didiagnosis adalah Diabetes; apakah metode diagnosis Diabetes jelas bersifat komputasional; apakah diagnosis Diabetes jelas hanya pada manusia. Pemilihan juga dilakukan berdasarkan judul studi dan abstrak.

Data berikut diekstraksi dari setiap publikasi: judul, penulis, referensi, database, jurnal, faktor keberhasilan kritis (yaitu faktor yang memiliki pengaruh positif pada pengembangan alat diagnosis Diabetes menggunakan metode komputasional), metodologi, target audiens atau populasi, deskripsi kualitas publikasi dan tahun. Kategorisasi data didasarkan pada beberapa metrik kinerja dari sampel publikasi final. Tinjauan menyeluruh dilakukan untuk mendapatkan daftar kategori untuk membantu dalam klasifikasi metrik kinerja. Daftar awal 12 kategori diidentifikasi dan selanjutnya digabungkan menjadi 7 faktor yang disajikan pada bagian 3.1. Pemilihan akhir sumber utama dari daftar makalah yang dipilih sebelumnya tidak hanya didasarkan pada membaca judul dan abstrak saja tetapi diperluas dengan membaca seluruh makalah.

### 2.2 Kriteria Evaluasi

Kinerja model dievaluasi menggunakan metrik berikut:

* Akurasi: Persentase prediksi yang benar dari total prediksi.
* Presisi: Proporsi prediksi positif yang benar dari total prediksi positif.
* Recall: Proporsi prediksi positif yang benar dari total kasus positif yang sebenarnya.
* F1-score: Rata-rata harmonis dari presisi dan recall, memberikan gambaran yang lebih baik tentang kinerja model ketika ada ketidakseimbangan kelas.

### 2.3 Model Komputasional yang Dianalisis

Model yang dianalisis dalam penelitian ini meliputi:

**Decision Tree**: Metode ini menggunakan struktur pohon untuk membuat keputusan berdasarkan fitur input. Setiap cabang pohon mewakili keputusan berdasarkan nilai fitur, dan setiap daun mewakili hasil akhir (diagnosis).

**K-Nearest Neighbors (KNN)**: Metode ini mengklasifikasikan data berdasarkan kedekatan dengan data lain. KNN menghitung jarak antara titik data yang tidak diketahui dan titik data yang sudah diketahui, kemudian mengklasifikasikan berdasarkan mayoritas kelas dari k tetangga terdekat.

**Naive Bayes**: Metode probabilistik ini menggunakan teorema Bayes dengan asumsi independensi antar fitur. Naive Bayes menghitung probabilitas setiap kelas berdasarkan fitur yang ada dan memilih kelas dengan probabilitas tertinggi.

**Logistic Regression**: Metode ini digunakan untuk memodelkan probabilitas dari dua kelas (diabetes atau tidak diabetes) berdasarkan fitur input. Logistic Regression menghasilkan fungsi logit yang dapat digunakan untuk memprediksi kelas.

**2.3 Kriteria Evaluasi**

Kinerja model dievaluasi menggunakan metrik berikut:

* Akurasi: Persentase prediksi yang benar dari total prediksi.
* Presisi: Proporsi prediksi positif yang benar dari total prediksi positif.
* Recall: Proporsi prediksi positif yang benar dari total kasus positif yang sebenarnya.
* F1-score: Rata-rata harmonis dari presisi dan recall, memberikan gambaran yang lebih baik tentang kinerja model ketika ada ketidakseimbangan kelas.

## 3. HASIL

### 3.1 Statistik Studi Terpilih

Dari 250 studi yang ditemukan, 50 artikel memenuhi kriteria inklusi. Hasil analisis menunjukkan bahwa Logistic Regression dan Decision Tree memiliki akurasi rata-rata di atas 85%, sedangkan KNN dan Naive Bayes memiliki akurasi rata-rata di bawah 80%. Tabel 1 menunjukkan ringkasan hasil dari masing-masing metode.

### 3.2 Perbandingan Kinerja

Tabel 1 menunjukkan perbandingan kinerja dari masing-masing metode:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Metode** | **Akurasi (%)** | **Presisi (%)** | **Recall (%)** | **F1-score (%)** |
| Decision Tree | 87 | 85 | 84 | 84.5 |
| K-Nearest Neighbors | 78 | 75 | 76 | 75.5 |
| Naive Bayes | 76 | 74 | 73 | 73.5 |
| Logistic Regression | 89 | 88 | 87 | 87.5 |

Tabel 2 Perbandingan Model Komputasi untuk Diagnosis Diabetes

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model Komputasi** | **Kekuatan** | **Kelemahan** | **Akurasi (%)** | **Kecepatan (detik)** | **Referensi** |
| Jaringan Saraf | Mampu menangkap hubungan non-linear | Memerlukan data besar untuk pelatihan | 85-95 | 2-5 | [1], [2], [3] |
| Logika Fuzzy | Fleksibel dalam menangani ketidakpastian | Hasil tergantung pada pemilihan fungsi keanggotaan | 75-85 | 1-3 | [4], [5] |
| Metode Bayesian | Memberikan probabilitas untuk diagnosis | Memerlukan asumsi yang mungkin tidak valid dalam konteks klinis | 80-90 | 3-6 | [6], [7], [8] |
| Algoritma Genetik | Baik untuk masalah kompleks dan non-linear | Akurasi dapat bervariasi tergantung pada parameter yang dipilih | 70-80 | 5-10 | [9], [10] |
| Sistem Ahli | Dapat meniru penalaran manusia | Bergantung pada pengetahuan yang diberikan oleh ahli | 75-85 | 2-4 | [11], [12] |

## 4. DISKUSI

Hasil penelitian menunjukkan bahwa Logistic Regression dan Decision Tree lebih efektif dalam mendiagnosis diabetes dibandingkan dengan KNN dan Naive Bayes. Logistic Regression memiliki keunggulan dalam hal interpretabilitas, memungkinkan dokter untuk memahami faktor-faktor yang mempengaruhi diagnosis. Di sisi lain, Decision Tree menawarkan visualisasi yang mudah dipahami, yang dapat membantu dalam pengambilan keputusan klinis. Namun, KNN dan Naive Bayes dapat menjadi pilihan yang baik dalam situasi tertentu, terutama ketika data tidak seimbang atau ketika kecepatan prediksi menjadi prioritas..

## 5. KESIMPULAN

Model komputasional yang digunakan dalam penelitian ini menunjukkan potensi yang signifikan dalam diagnosis diabetes. Logistic Regression dan Decision Tree adalah metode yang paling efektif, tetapi penelitian lebih lanjut diperlukan untuk meningkatkan akurasi dan kecepatan diagnosis. Pengembangan model yang lebih kompleks dan penggunaan teknik ensemble dapat menjadi arah penelitian selanjutnya. Selain itu, integrasi model ini dengan sistem kesehatan yang ada dapat membantu dalam pengelolaan diabetes secara lebih efektif.

## 6. DAFTAR PUSTAKA

Cahyani, Q. R., Finandi, M. J., Rianti, J., Arianti, D. L., Dwi, A., Putra, P., & Artikel, G. (2022). Prediksi Risiko Penyakit Diabetes menggunakan Algoritma Regresi Logistik Diabetes Risk Prediction using Logistic Regression Algorithm Article Info ABSTRAK. JOMLAI J. Mach. Learn. Artif. Intell, 1(2), 2828-9099.

Putri, E. F. (2024). Analisis Faktor-Faktor yang Memengaruhi Status Diabetes Mellitus pada Pra Lansia dan Lansia di Indonesia Menggunakan Model Regresi Logistik Biner. Statistika, 24(1), 54-64.

Gunawan, S., Astuti, R., Prihartono, W., & Hamonangan, R. (2025). PREDIKSI DIABETES MELLITUS TIPE 2 DENGAN ALGORITMA LOGISTIC REGRESSION UNTUK PENDETEKSIAN DINI. Jurnal Informatika dan Teknik Elektro Terapan, 13(1).

Mutaqin, Z., Rozikin, C., & Tomo, Y. A. (2024). KLASIFIKASI PENYAKIT DIABETES MENGGUNAKAN ALGORITMA LOGISTIC REGRESSION. Jurnal Sistem dan Teknologi Informasi (JSTI), 6(3).

Tyasnurita, R., & Pamungkas, A. Y. M. (2020). Deteksi Diabetik Retinopati menggunakan Regresi Logistik. ILKOM Jurnal Ilmiah, 12(2), 130-135.

Andriani, A. (2013). Sistem prediksi penyakit diabetes berbasis decision tree. Bianglala Informatika, 1(1).

Hana, F. M. (2020). Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision Tree C4. 5. Jurnal SISKOM-KB (Sistem Komputer dan Kecerdasan Buatan), 4(1), 32-39.

Afifuddin, A., & Hakim, L. (2023). Deteksi Penyakit Diabetes Mellitus Menggunakan Algoritma Decision Tree Model Arsitektur C4. 5. Krisnadana Journal, 3(1), 25-33.

Safitri, L., & Fatah, Z. (2024). Implementasi Prediksi Penyakit Diabetes Menggunakan Metode Decision Tree. JUSIFOR: Jurnal Sistem Informasi dan Informatika, 3(2), 125-132.

Aditya, M. F., Pramuntadi, A., Wijaya, D. P., & Wicaksono, Y. (2024). Implementasi Metode Decision Tree pada Prediksi Penyakit Diabetes Melitus Tipe 2: Implementation of Decision Tree Method for Diabetes Mellitus Type 2 Prediction. MALCOM: Indonesian Journal of Machine Learning and Computer Science, 4(3), 1104-1110.

Devian, D., Sabrina, P. N., & Komarudin, A. (2024). PREDIKSI PENYAKIT DIABETES DENGAN METODE K-NEAREST NEIGHBOR (KNN) DAN SELEKSI FITUR INFORMATION GAIN. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *8*(6), 11320-11326.

Huwaidi, F., Taufikurachman, H., & Rilwanu, M. F. N. Penerapan Algoritma K-Nearest Neighbor untuk Mendeteksi Diabetes Berbasis Web Application.

Fasnuari, H. A. D., Yuana, H., & Chulkamdi, M. T. (2022). Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Penyakit Diabetes Melitus. *Antivirus: Jurnal Ilmiah Teknik Informatika*, *16*(2), 133-142.

Wijaya, M. R., Rasywir, E., Meisak, D., & Pratama, Y. (2022). Implementasi Algoritma K-Nearst Neighbor Untuk Memprediksi Penyakit Diabetes. *Jurnal Informatika Dan Rekayasa Komputer (JAKAKOM)*, *2*(2), 231-239.

Arti, P. P., Yanti, I., & Pauzan, M. (2024). ELECTRONIC NOSE UNTUK SKRINING PENYAKIT DIABETES MELLITUS MENGGUNAKAN K-NEAREST NEIGHBOR DAN LARIK SENSOR. *Jurnal Teknologi Informasi dan Ilmu Komputer*, *11*(2), 429-436.

Rinanda, P. D., Delvika, B., Nurhidayarnis, S., Abror, N., & Hidayat, A. (2022). Perbandingan Klasifikasi Antara Naive Bayes dan K-Nearest Neighbor Terhadap Resiko Diabetes pada Ibu Hamil: comparison of classification between Naive Bayes and k-nearest neighbor on diabetes risk in pregnant women. *Malcom: Indonesian Journal of Machine Learning and Computer Science*, *2*(2), 68-75.

Fitriyani, F. (2021). Prediksi Diabetes Menggunakan Algoritma Naive Bayes dan Greedy Forward Selection. *Jurnal Nasional Teknologi dan Sistem Informasi*, *7*(2), 61-69.

Hamzah, I., & Sitorus, Z. (2024). Analisa Classification Decision Tree C45 dan Naïve Bayes Pada Indikasi Penyakit Diabetes Menggunakan Rapid Miner. *Jurnal Nasional Teknologi Komputer*, *4*(1), 25-33.

Widodo, Y. B., Anggraeini, S. A., & Sutabri, T. (2021). Perancangan Sistem Pakar Diagnosis Penyakit Diabetes Berbasis Web Menggunakan Algoritma Naive Bayes. *J. Teknol. Inform. dan Komput*, *7*(1), 112-123.

Agustin, A. V., & Voutama, A. (2023). Implementasi Data Mining Klasifikasi Penyakit Diabetes Pada Perempuan Menggunakan Naïve Bayes. JATI (Jurnal Mahasiswa Teknik Informatika), 7(2), 1002-1007.