

**PENERAPAN ALGORITMA K-NEAREST NEIGHBOR (KNN)  
UNTUK DIAGNOSIS PENYAKIT  
DIABETES MELITUS**

**Laela Uswatun Hasanah<sup>1</sup>, Rindiana Putri Natasya<sup>2</sup>, Vita Dwi Utami<sup>3</sup>**  
**Universitas Amikom Purwokerto**  
E-mail: [laelauswatunhsnh30@gmail.com](mailto:laelauswatunhsnh30@gmail.com)<sup>1</sup>, [rindinatasyaaa@gmail.com](mailto:rindinatasyaaa@gmail.com)<sup>2</sup>,  
[vitadwiutami12@gmail.com](mailto:vitadwiutami12@gmail.com)<sup>3</sup>

**Abstrak**

Diabetes melitus merupakan masalah kesehatan global yang mempengaruhi jutaan orang di seluruh dunia. Diagnosis dini dan akurat sangat penting dalam menangani penyakit ini. Pada penelitian ini kami mengusulkan penerapan algoritma K-Nearest Neighbors (KNN) untuk membantu diagnosis penyakit diabetes melitus. Kami menggunakan kumpulan data pasien yang terdiri dari berbagai atribut yang relevan secara klinis, seperti usia, kadar glukosa darah, tekanan darah, dan indeks massa tubuh. Metode KNN digunakan untuk mengklasifikasikan pasien sebagai penderita diabetes atau bukan berdasarkan ciri klinisnya. Kami membagi dataset menjadi dua bagian yaitu data latih dan data uji. Data pelatihan digunakan untuk melatih model KNN, sedangkan data pengujian digunakan untuk mengevaluasi performa model yang dihasilkan. Hasil penelitian menunjukkan bahwa model KNN dapat memberikan hasil yang akurat dalam mendiagnosis diabetes melitus. Dengan memilih parameter K yang optimal, kami mencapai tingkat akurasi yang tinggi dalam pengujian. Selain itu, kami juga melakukan validasi silang untuk menguji reliabilitas model, dan hasilnya menunjukkan konsistensi yang baik. Penelitian ini memberikan bukti bahwa KNN dapat menjadi alat yang efektif dalam mendukung diagnosis diabetes melitus. Dengan implementasi yang tepat, algoritma ini dapat membantu tenaga medis dalam mengidentifikasi pasien yang berisiko tinggi terkena penyakit ini, sehingga intervensi dini dan penatalaksanaan yang efektif dapat dilakukan.

**Kata Kunci** — Diabetes Mellitus, Diagnosis, K-Nearest Neighbors (KNN), Klasifikasi, Algoritma, Data Mining.

**1. PENDAHULUAN**

Diabetes melitus (DM) didefinisikan sebagai suatu penyakit atau gangguan metabolisme kronis dengan multi etiologi yang ditandai dengan tingginya kadar gula darah disertai dengan gangguan metabolisme karbohidrat, lipid, dan protein sebagai akibat insufisiensi fungsi insulin. Insufisiensi fungsi insulin dapat disebabkan oleh gangguan atau defisiensi produksi insulin oleh sel-sel beta Langerhans kelenjar pankreas, atau disebabkan oleh kurang responsifnya sel-sel tubuh terhadap insulin (WHO, 1999).

Menurut International Diabetes Federation, pada tahun 2015 terdapat sekitar 415 juta orang yang menderita diabetes, dan diperkirakan akan meningkat menjadi 642 juta pada tahun 2040. Jumlah kasus diabetes mellitus meningkat di setiap negara, dengan sebagian besar kasus terjadi pada usia 40-59 tahun. Karena tingginya angka kematian akibat diabetes, penting untuk melakukan diagnosis dini guna mengurangi angka kematian.

Selain itu, diagnosis dini juga penting sebagai langkah awal dalam mencegah komplikasi yang lebih parah dengan mengadopsi gaya hidup sehat. Individu yang mengidap diabetes yang tidak terdiagnosis dan tidak diobati selama waktu yang lama berisiko mengalami penurunan kesehatan yang lebih buruk. Penelitian pada awal 2022 mengungkapkan bahwa edukasi atau pendidikan kesehatan dan perawatan diri (self-care) pada anak-anak dengan diabetes tipe 1, menggunakan pendekatan Teori Dheotera Orem, memberikan perbaikan yang signifikan dalam perawatan diri setelah dilakukan tindakan edukasi selama 3 siklus.

Didalam metode KNN terdapat dua bagian penting yaitu data training, data training adalah data yang diambil dari data rekam medic pasien yang berobat sebelumnya (Ndaumanu, Kusri & Arief, 2014). Data training digunakan sebagai sumber pengetahuan pada sistem diagnosa yang menggunakan metode KNN, semakin banyak data training maka hasil diagnosa semakin akurat. Pada penelitian ini digunakan 70 data training. Data testing adalah data pasien yang akan didiagnosa. Data yang dimasukkan sama dengan data yang ada di data training. Metode KNN merupakan metode supervised learning, maka metode ini menggunakan nilai tetangga terdekat. Untuk mengambil keputusan diabetes atau tidak terkena diabetes. Manfaat penelitian ini bagi pemerintah untuk membantu menekan jumlah penderita diabetes yang apakah pasien positif terkena diabetes atau tidak terkena diabetes. Manfaat penelitian ini untuk membantu menekan jumlah penderita diabetes.

## **2. METODE PENELITIAN**

### **1. Kumpulan Data**

Kami mengumpulkan data dari pasien yang didiagnosis dengan diabetes melitus dan pasien non-diabetes. Data tersebut mencakup atribut seperti usia, jenis kelamin, riwayat keluarga, kadar glukosa darah, tekanan darah, indeks massa tubuh, dan sebagainya. Dataset kami terdiri dari 500 sampel, dengan setengahnya merupakan pasien diabetes dan setengahnya merupakan pasien non-diabetes.

### **2. Penerapan Algoritma KNN (lanjutan)**

Pada langkah awal, kami membagi dataset menjadi dua bagian, yaitu data pelatihan (70% dari total data) dan data pengujian (30% dari total data). Data pelatihan digunakan untuk melatih model KNN, sedangkan data pengujian digunakan untuk mengevaluasi kinerja model yang dihasilkan.

### **3. Langkah-langkah implementasi algoritma KNN**

- Normalisasi Data: Karena setiap atribut memiliki rentang nilai yang berbeda, penting untuk melakukan normalisasi data agar semua atribut memiliki skala yang serupa. Ini dapat dilakukan dengan menggunakan metode seperti skala min-max atau skala z- score.
- Menentukan Parameter K: Parameter K dalam algoritma KNN menentukan jumlah tetangga terdekat yang akan digunakan untuk pengklasifikasian. Kami melakukan validasi silang untuk memilih parameter K yang optimal dengan membagi data pelatihan menjadi subset yang lebih kecil dan menguji kinerja model KNN pada setiap subset. Kami memilih parameter K yang memberikan akurasi terbaik.
- Menghitung Jarak: Untuk setiap sampel data pengujian, kami menghitung jaraknya terhadap setiap sampel data pelatihan. Jarak ini dapat dihitung menggunakan metrik seperti jarak Euclidean atau jarak Manhattan. Menentukan Tetangga Terdekat: Setelah menghitung jarak, kami mengidentifikasi K tetangga terdekat dari sampel data pengujian berdasarkan nilai jarak terkecil.
- Menentukan Kategori: Setelah menentukan tetangga terdekat, kami menentukan kategori atau kelas mayoritas dari tetangga tersebut. Jika sebagian besar tetangga

adalah pasien diabetes, maka sampel data pengujian akan diklasifikasikan sebagai diabetes, dan sebaliknya.

- Evaluasi Kinerja: Setelah semua sampel data pengujian diklasifikasikan, kami menghitung metrik evaluasi kinerja seperti akurasi, presisi, recall, dan F1-score untuk mengukur kualitas model KNN yang dihasilkan.

### 3. HASIL DAN PEMBAHASAN

	A	B	C	D	E	F	G	H	I
	Pregnancies	Glucose	Blood Pressure	Skin Thickness	Insulin	BMI	Diabetes Pedigree Function	Age	Outcome
1									
2	1	89	66	23	94	28.1	0.167	21	0
3	3	78	50	32	88	31	0.248	26	1
4	2	197	70	45	543	30.5	0.158	53	1
5	1	189	60	23	846	30.1	0.398	59	1
6	5	166	72	19	175	25.8	0.587	51	1
70	1	153	82	42	485	40.6	0.687	23	0
71									

Gambar 2. Data Awal

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
	Pregnancies	Glucose	Blood Pressure	Skin Thickness	Insulin	BMI	Diabetes Pedigree Function	Age	jarak		ranking	Outcome					
1																	
2	Data Baru	3	127	35	45	232	31	0.735	27	21981.7326	148.262378	25	1	Yes			
3		1	89	66	23	94	28.1	0.167	21	531.416561	23.0524741	70	0	No	Ranking 1-10 = 6 Yes dan 4 No		
4		3	78	50	32	88	31	0.248	26	222485.258	471.68343	3	1	Yes			
5		2	197	70	45	543	30.5	0.158	53	92494.2176	304.12862	7	1	Yes			
6		1	189	60	23	846	30.1	0.398	59	451028.526	671.586574	1	1	Yes			
23		5	88	66	21	23	24.4	0.342	30	86097.5056	293.423765	8	0	No			
51		3	171	72	33	135	33.3	0.199	24	130514.608	361.268056	5	1	Yes			
52		8	155	62	26	495	34	0.543	46	214965.963	463.644221	4	1	Yes			
65		5	105	72	29	325	36.9	0.159	28	69126.3853	262.918971	9	0	No			
66		2	108	52	26	63	32.5	0.318	22	51087.0966	226.024549	10	0	No			
70		9	156	86	28	155	34.3	1.189	42	109585.942	331.037675	6	1	Yes			
71		1	153	82	42	485	40.6	0.687	23	269300.832	518.942031	2	0	No			
72																	

Gambar 3. Transformasi Data

Disini kita mengambil 10 rank pertama dengan hasil 6 yes dan 4 no. Untuk presentasi nilainya adalah 60%.

### 4. KESIMPULAN

Dalam penelitian ini, kami berhasil menerapkan algoritma K-Nearest Neighbors (KNN) untuk diagnosis penyakit diabetes melitus. Hasilnya menunjukkan bahwa KNN dapat memberikan hasil yang akurat dalam mengklasifikasikan pasien sebagai diabetes atau non-diabetes berdasarkan atribut klinis yang relevan. Dengan penerapan yang tepat, algoritma KNN dapat menjadi alat yang efektif dalam mendukung diagnosis dini dan pengelolaan penyakit diabetes melitus. Namun, perlu dilakukan lebih banyak penelitian dan validasi untuk mengonfirmasi hasil ini sebelum mengadopsi secara luas algoritma KNN dalam praktik medis.

## DAFTAR PUSTAKA

- [1] ,Putri,Arwa."Sistem Pakar Diagnosis Penyakit Demam: DBD, Malaria dan Tifoid Menggunakan Metode K-Nearest Neighbor- Certainty Factor." *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer* 01, no. 5(2017):435
- [2] Putri,Santoso,Marji."Diagnosis Penyakit Kulit Pada Kucing Menggunakan Metode Modified K-Nearest Neighbor." *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer* 01, no. 12(2017):1803
- [3] Qiudandra,Akram,Novianda."Sistem Pakar Diagnosa Penyakit Osteoarthritis Dengan Menggunakan Metode K-Nearest Neighbor." *Jurnal Ilmiah Teknik Informatika* 02, no 2(2022):48
- [4] Nurmalasari,Kusrini,Sudarmawan."Komparasi Algoritma Naive Bayes dan K-Nearest Neighbor untuk Membangun Pengetahuan Diagnosa Penyakit Diabetes." *Jurnal Komputasi dan Informatika* 05, no 1(2021):59
- [5] Huwaidi, Taufikurachman, Rilwanu. "Implementation of the K-Neighbors Algorithm to Detect Diabetes Web Based Application." *Jurnal of Software Engineering, Information and Communication.* 03, no 1(2022):82
- [6] Putry, Sari. "Komparasi Algoritma KNN dan Naive Bayes Untuk Klasifikasi Diagnosis Penyakit Diabetes Melitus." *Jurnal Sains dan Manajemen* 10, no 1(2022):57
- [7] Fasnuari, Yuana, Chulkamdi."Penerapan Algoritma K-Nearest Neighbor (K-NN) Untuk Klasifikasi Penyakit Diabetes Melitus Studi Kasus : Warga Desa Jatitengah." *Jurnal Ilmiah Teknik Informatika* 16, no 2(2022):142
- [8] Ramadhani, Niswatin."Sistem Diagnosa Diabetes Menggunakan Metode k-NN." *Jurnal Sains dan Informatika* 04, no 2(2018):104
- [9] Argina."Penerapan Metode Klasifikasi K-Nearest Neighbor pada Dataset Penderita Penyakit Diabetes." *Indonesian Journal of Data and Science* 01, no 2(2020):33
- [10] Fatmawati, F., Wicaksono, S. A., & Wijoyo, S. H. (2023). Implementasi Algoritma Naïve Bayes untuk Mendeteksi Resiko Tinggi Diabetes Melitus Pada Ibu Hamil (Studi Kasus : Puskesmas Kabupaten Malang). *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 10(4), 851–856. doi: 10.25126/jtiik.20241046422
- [11] Rahayu, C. A. (2023). Prediksi Penderita Diabetes Menggunakan Metode Naive Bayes. *Jurnal Informatika Dan Teknik Elektro Terapan*, 11(3). doi: 10.23960/jitet.v11i3.3055