Diterima : 13 Juni 2025
Disetujui : 25 Juni 2025
Hal : 144-151

Al-Aqlu: Jurnal Matematika, Teknik dan Sains
https://jurnal.yalamqa.com/index.php/aqlu
e-ISSN : 2985-4369
Vol. 3 , No. 2, Juli 2025

PERBANDINGAN KINERJA METODE K-NEAREST NEIGHBOR (KNN), RANDOM FOREST, DAN DECISION TREE DALAM MEMPREDIKSI DIABETES

[Comparing the Accuracy of K-Nearest Neighbour (KNN), Random Forest, and Decision Tree Methods in Predicting Diabetes]

Sherly Yulianty[1)], Mohamad Khoirun Najib[2)]*

Sekolah Sains Data, Matematika dan Informatika, IPB University

mkhoirun@apps.ipb.ac.id (corresponding author)

ABSTRAK

Diabetes merupakan suatu penyakit dengan jumlah penderita yang terus bertambah dan menjadi penyebab kematian dari 1.5 juta manusia di dunia pada tahun 2019. Diperlukan suatu penanganan penyakit diabetes, salah satunya dengan melakukan prediksi penderita diabetes. Metode *K-Nearest Neighbor* (KNN), *Random Forest*, dan *Decision Tree* merupakan beberapa metode yang dapat digunakan untuk melakukan prediksi klasifikasi diabetes. Penelitian ini bertujuan membandingkan kinerja metode KNN, *Random Forest*, dan *Decision Tree* berdasarkan akurasi dan waktu komputasinya. Data yang digunakan pada penelitian ini yaitu *Pregnancies*, *Glucose*, Insulin, *Body Mass Index* (BMI), dan *Age* sebagai peubah bebas serta *Outcome* sebagai peubah terikat. Hasil penelitian data yang belum dinormalisasi dengan Min-Max menunjukkan metode KNN memiliki waktu komputasi yang lebih cepat dibandingkan dua metode lainnya, sedangkan berdasarkan nilai akurasinya metode *Decision Tree* memiliki nilai yang lebih tinggi dibandingkan dua metode lainnya. Selanjutnya pada data yang telah dinormalisasi Min-Max menunjukkan penurunan nilai akurasi pada metode *Decision Tree* dan *Random Forest*, sedangkan nilai akurasi metode KNN mengalami peningkatan. Oleh karena itu, perlakuan normalisasi Min-Max lebih baik digunakan untuk metode KNN.

Kata kunci: *Diabetes; K-Nearest Neighbor; Random Forest; Decision Tree*

ABSTRACT

Diabetes is a disease with a growing number of sufferers and is the cause of death of 1.5 million people in the world in 2019. A treatment for diabetes is needed, one of which is by predicting diabetics. The K-Nearest Neighbour (KNN), Random Forest, and Decision Tree methods are some methods that can be used to predict diabetes classification. This research aims to compare the performance of KNN, Random Forest, and Decision Tree methods based on accuracy and computation time. The data used in this study are Pregnancies, Glucose, Insulin, Body Mass Index (BMI), and Age as independent variables and Outcome as a dependent variable. The results of research on data that has not been normalised with Min-Max show that the KNN method has a faster computation time than the other two methods, while based on the accuracy value the Decision Tree method has a higher value than the other two methods. Furthermore, the Min-Max normalised data shows a decrease in the accuracy value of the Decision Tree and Random Forest methods, while the accuracy value of the KNN method has increased. Therefore, the Min-Max normalisation treatment is better used for the KNN method.

Keywords: *Diabetes; K-Nearest Neighbor; Random Forest; Decision Tree*

PENDAHULUAN

Saat ini, teknologi telah berkembang dengan pesat yang turut berpengaruh dalam keberlangsungan hidup manusia. Tekonologi berperan aktif dalam keberlangsungan hidup manusia di berbagai bidang, misalnya bidang pendidikan, bisnis, kesehatan, informasi, komunikasi, transportasi, energi, dan masih banyak lainnya. Secara khusus pada bidang kesehatan, berbagai jenis

penerapan teknologi telah memberikan perannya yang di antaranya teknologi informasi, komunikasi, medis, dan lainnya. *Machine learning* sebagai bagian dari teknologi informasi telah terbukti memiliki banyak manfaat di bidang kesehatan (Telaumbanua et al., 2019). Penerapan *machine learning* yang telah banyak dilakukan yaitu misalnya dalam melakukan klasifikasi pasien dalam diagnosa suatu penyakit. Contoh penerapan *machine learning* dalam bidang kesehatan misalnya adalah penggunaan metode *Decision Tree* yang telah terbukti memiliki akurasi yang sangat baik dalam melakukan prediksi klasifikasi tingkat suatu penyakit (Wardhana et al., 2023). Selain itu, terdapat metode lainnya yang dapat digunakan dalam proses klasifikasi penyakit yang di antaranya adalah *K-Nearest Neighbor, Random Forest*, dan lain sebagainya.

Diabates merupakan suatu penyakit dengan jumlah penderitanya yang terus bertambah. Diabetes menjadi penyebab dari 1.5 juta kematian yang ada pada tahun 2019 (WHO, 2023). Oleh karena itu, perlu dilakukan pencegahan dan penanganan terkait dengan penyakit diabetes. Salah satu pencegahan dan penanganan yang dapat dilakukan yaitu dengan melakukan prediksi pendiagnosaan diabetes pada pasien. Pada penelitian ini akan dilakukan perbandingan prediksi akurasi klasifikasi penyakit diabetes dengan metode *K-Nearest Neighbor (KNN), Random Forest,* dan *Decision Tree.*

Penelitian terkait dengan prediksi diagnosa penyakit dengan model klasifikasi telah banyak dilakukan. Pada penelitiannya, Aprilliandhika dan Abdulloh (2024) membandingkan kinerja KNN dan *Support Vector Machine* (SVM) dalam prediksi penyakit stroke. Berdasarkan penelitian tersebut diketahui bahwa model KNN lebih baik digunakan untuk prediksi penyakit stroke dibandingkan SVM. Selanjutnya, Aziz et al. (2023) melakukan penelitian terkait dengan prediksi penyakit jantung dengan model klasifikasi yang berbasis *Decision Tree*.

Pada penelitian terdahulu yang telah dijelaskan sebelumnya, metode klasifikasi KNN, *Random Forest*, dan *Decision Tree* dilakukan secara terpisah dengan menggunakan data yang berbeda pada setiap penelitian. Selain itu pada penelitian terdahulu, perbandingan model klasifikasi pada dataset yang sama dilakukan pada dua model klasifikasi. Sedangkan pada penelitian ini akan dilakukan perbandingan kinerja model klasifikasi yaitu KNN, *Random Forest*, dan *Decision Tree* pada dataset yang sama. Pemilihan model terbaik dilakukan dengan melihat tingkat akurasi model klasifikasi dalam melakukan prediksinya.

METODE PENELITIAN

Data yang digunakan pada penelitian ini adalah data klasifikasi pasien diabetes yang diperoleh dari Kaggle. Data tersebut diperoleh dari 768 pasien yang diamati. Peubah terikat yang digunakan pada penelitian ini adalah diagnosa pasien terkait penyakit diabetes (*Outcome*). Selanjutnya untuk peubah bebas yang digunakan pada penelitian ini terdiri atas *Pregnancies, Glucose, Insulin, Body Mass Index* (BMI), dan *Age*. Data yang digunakan pada penelitian ini dibagi menjadi data *training* dan *testing*. Pembagian data tersebut dilakukan dengan menggunakan data baru yang telah dilakukan penanganan ketidakseimbangan data yang telah dilakukan sebelumnya. Data *training* yang digunakan yaitu 80% dari total data yang digunakan untuk membangun model dan sisanya digunakan sebagai data *testing*.

K-Nearest Neighbor (KNN)

KNN merupakan suatu metode klasifikasi yang paling umum digunakan. KNN merupakan suatu metode klasifikasi yang dibangun pada data *training* berdasarkan jarak yang paling dekat dengan objek berdasarkan nilai k (Setianto et al., 2018, Hawari et al., 2024). Metode KNN pertama kali diperkenalkan pada tahun 1951 oleh Fix dan Hodges. Metode KNN memiliki kelebihan yang di antaranya adalah memiliki tingkat akurasi yang tinggi dan tidak memiliki asumsi khusus yang harus dipenuhi oleh data yang digunakan (Pratama et al., 2022). KNN merupakan suatu metode *supervised classification* (Gharehbaghi, 2023). K tetangga terdekat dari data *testing,* $Y, (Y \in \mathbb{R}^n)$ ditemukan dengan jarak Euclidean yang dirumuskan seperti berikut,

$$D_i = (Y - X_i)^T * (Y - X_i)\,.$$

Algoritma 1 Classification of a testing sample, Y, based on KNN method

Procedure KNN$(\{V\}, Y, K)$
  Calculate$\{D_i(Y, \{X_i\})\}$
  $S = Sort(\{D_i\}, Descending)$
  $S_T = \{S_i : i = 1, \dots, K\}$
  $I \leftarrow \arg\{S_T\}$
  $q \leftarrow \arg\max_i\{Q_i(I)\}$
  return q
end procedure

Random Forest

Random Forest merupakan suatu algoritma *machine learning* yang digunakan untuk proses klasifikasi. *Random Forest* merupakan suatu kumpulan pohon klasifikasi yang diperoleh dari *sampling bootstrip* data (Chairunisa et al., 2020). Rumus menentukan pohon keputusan, dituliskan seperti berikut,

$$Entrophy\ (S) = \sum_{i=1}^{n} (-P_i \times \log 2\ P_i)$$

$$Gain(S, A) = Entrophy\ (S) - \sum_{i=1}^{n} \left(\frac{|S_i|}{|S|} \times Entrophy(S_i) \right)$$

dengan S merupakan kumpulan data yang diamati, P_i merupakan persentase dari setiap bagian S_i terhadap total seluruh himpunan S, dan $|S|$ merupakan banyaknya total kasus dalam himpunan S (Handayani et al., 2024). Selanjutnya algoritma *Random Forest* dapat dituliskan seperti pada Algoritma 2 berikut (Kozak, 2019).

Algoritma 2 Random forest algorithm

```
ensemble = NULL;
for number_of_classifiers do
  // Construction of a decision tree classifier
  data_set_classifier = choose_objects(data_set); // bootstrap aggregation
  new_classifier = NULL;
  while incomplete_decision_tree
    attributes = create_subset_of_attributes(√k from all atributes);
    division = select_next_division(data_set_classifier, attributes);
    new_classifier.add(division);
  endWhile
  ensemble.add(new_classifier);
endFor
result = ensemble;
```

Decision Tree

Decision Tree merupakan suatu metode yang digunakan untuk proses klasifikasi yang berbasis pohon. Terdapat beberapa algoritma yang termasuk kedalam *Decision Tree* yang di antaranya ID3, C4.5, dan CART (Solahuddin et al., 2023). *Decision Tree* memiliki kelebihan yang menjadikannya banyak disukai dalam pemodelan untuk proses klasifikasi. *Decision Tree* memiliki kelebihan yang mampu menangani data kompleks dan beragam, serta memberikan model yang tidak sulit dalam diinterpretasikan (Rahman et al., 2024). Struktur *decision tree* mirip dengan *tree with a root node, a left subtree, and right subtree*. Node daun pada pohon mewakili label kelas. Busur dari satu node ke node lainnya menunjukkan kondisi pada atribut (Alifah et al., 2024).

Decision Tree $T(S)$ dapat digunakan dalam proses klasifikasi dengan mengikuti persamaan berikut (dengan D merupakan notasi distribusi),

$$\epsilon(T(S), D) = \sum_{(x,y)\in U} D(x,y) \cdot L\big(y, T(S)(x)\big),$$

dengan $L\big(y, T(S)(x)\big)$ diberikan pada fungsi berikut,

$$L\big(y, T(S)(x)\big) = \begin{cases} 0, & \text{jika } y = T(S)(x) \\ 1, & \text{jika } y \neq T(S)(x), \end{cases}$$

untuk $T(S)$ merupakan pohon keputusan T yang dibangun berdasarkan himpunan *training* S, $T(S)(x)$ merupakan keputusan untuk peubah x yang ditentukan oleh atribut kondisionalnya, dan U merupakan himpunan nilai yang mungkin untuk setiap atribut (Kozak, 2019). Algortima *Decision Tree* dapat dilihat pada Algoritma 3 berikut.

Algoritma 3 ***Pseudo-code of the ACDT algorithm***

```
pheromone = initialization_pheromone_trail();
for number_of_iterations do
 best_tree = NULL;
 for number_of_ants do
   //build the decision tree
   new_tree = null;
   while (stop_condition_is_not_fulfilled)
     heuristic = calculate_the_heuristic_function();
     p = calc_the_choossing_probability(pheromone, heuristic);
     //choose the test in the node (roulette wheel)
     new_tree⟶test = roulette_wheel(p_{m,m_{L(i,j)}}(t));
   endWhile
   pruning(new_tree);
   assessment_of_the_tree_quality(new_tree);
   if new_tree is_higher_quality_than bbest_tree then
     best_tree = new_tree;
   endIf
 endFor
 update_pheromone_trail(best_tree, pheromone);
 if best_tree is_higher_quality_than best_constructed_tree then
     best_constructed_tree = best_tree;
 endIf
endFor
result = best_constructed_tree;
```

Akurasi

Akurasi digunakan untuk mengukur kinerja model dalam melakukan klasifikasi dengan benar. Umumnya, nilai akurasi diberikan dalam bentuk persentase. Apabila nilai akurasi mendekati 1 (100%), maka model dapat melakukan prediksi klasifikasi dengan sangat baik. Nilai akurasi dapat dihitung dengan menggunakan rumus yang ditunjukkan pada persamaan berikut (Maskuri et al., 2022),

$$\text{Akurasi} = \frac{\text{Jumlah prediksi data benar}}{\text{Jumlah data } testing} \times 100\%.$$

Tahapan

Penelitian dilakukan dengan mengikuti langkah-langkah penelitian sebagai berikut:

1. *Merumuskan Masalah*. Penelitian ini diawali dengan merumuskan masalah yang selanjutnya menjadi dasar dan arah penelitian. Pada penelitian ini, peneliti tertarik untuk melakukan penelitian terkait dengan klasifikasi diagnosa penyakit diabetes berdasarkan dengan faktor-faktor penyebab yang dipilih peneliti dalam penelitian ini.

2. *Mengumpulkan Data*. Pada tahap ini, peneliti melakukan pengumpulan data yang dibutuhkan untuk masalah yang telah dirumuskan sebelumnya.
3. *Melakukan Eksplorasi Data*. Pada tahap ini, peneliti melakukan eksplorasi data untuk melihat karakteristik data. Selanjutnya peneliti melakukan penyaringan data secara spesifik untuk dapat digunakan pada penelitian ini.
4. *Menentukan Metode*. Pada tahap ini, peneliti melakukan penentuan metode yang sesuai dengan data yang digunakan. Metode-metode yang sesuai dengan suatu data dapat ditentukan pada bahasa pemrograman Julia dengan menggunakan package MLJ. Pada penelitian ini, peneliti memilih untuk menggunakan tiga metode yang di antaranya adalah KNN, Random Forest, dan Decision Tree.
5. *Melakukan Studi Literatur*. Pada tahap ini, peneliti melakukan studi literatur untuk memhami metode-metode yang digunakan pada penelitian ini. Peneliti mempelajari metode-metode yang digunakan melalui jurnal dan buku yang berkaitan yang telah tersebar luas di internet.
6. *Mengolah Data.* Data yang telah dikumpulkan dan disaring sebelumnya untuk memperoleh data-data yang paling sesuai dalam masalah yang diangkat, selanjutnya dilakukan pengolahan data dengan menggunakan metode-metode yang telah dipilih sebelumnya.
7. *Menyusun Makalah Penelitian.* Data yang telah selesai diolah, selanjutnya disajikan menjadi suatu tulisan yang disusun dalam suatu makalah. Dalam suatu penelitian, hasil penelitian perlu dituliskan dan dibagikan sebagai suatu informasi kepada pembaca dan dapat diimanfaatkan sebagaimana mestinya.
8. *Menarik Kesimpulan.* Penelitian ini diakhiri dengan penarikan kesimpulan atas penelitian yang dilakukan. Berdasarkan hasil kesimpulan dapat dilihat apakah terdapat kekurangan yang perlu dilakukan perbaikan pada penelitian ini.

HASIL DAN PEMBAHASAN

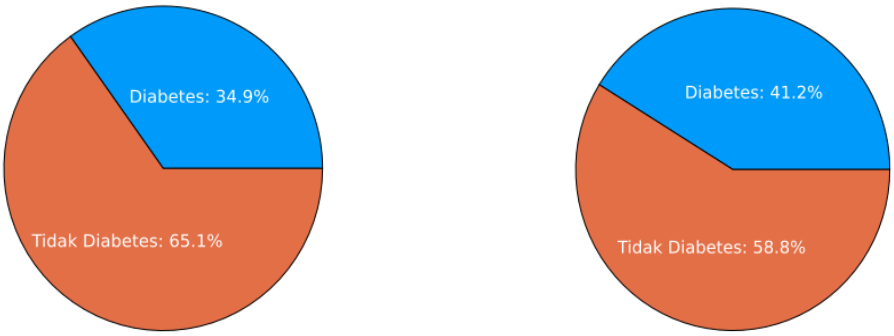
Eksplorasi Data

Sebelum mengolah data, penting untuk melakukan pengecekan missing data pada data yang digunakan. Banyaknya missing data pada data penelitian ini disajikan pada Tabel 1 berikut.

Tabel 1. Deteksi *Missing Data* Pada Setiap Peubah

Peubah	Banyaknya *missing data*
Pregnancies	0
Glucose	0
Insulin	0
Body Mass Index (BMI)	0
Age	0
Outcome	0

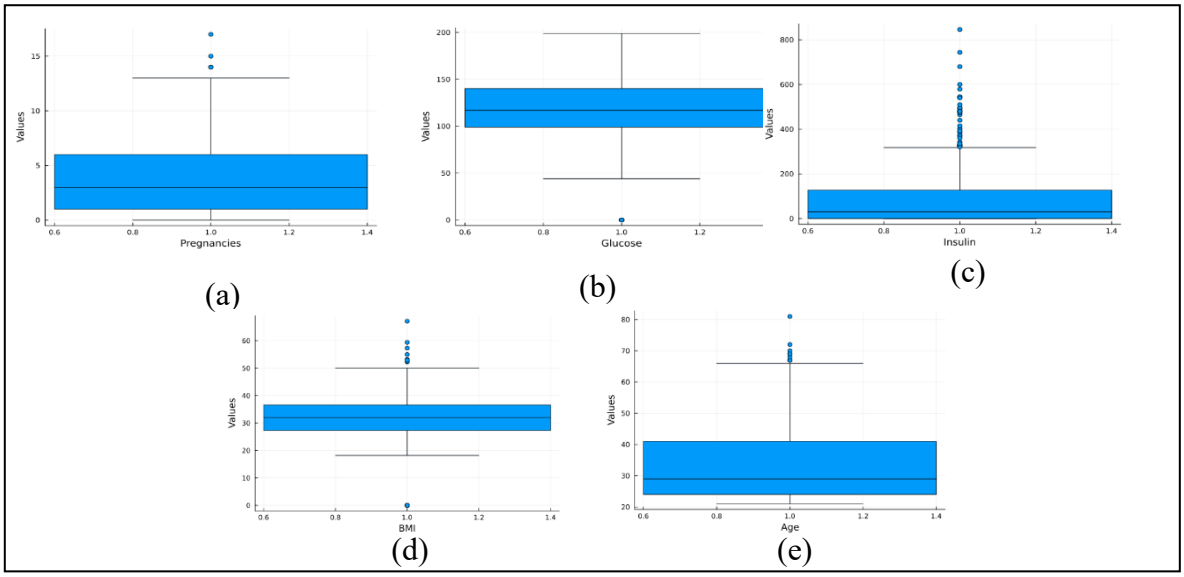
Berdasarkan Tabel 1, peubah-peubah yang digunakan pada penelitian ini tidak terdapat *missing data,* sehingga tingkat bias pada model yang terbentuk sekecil mungkin. Akan tetapi, pada data penelitian yang digunakan, terdapat ketidakseimbangan data antara diagnosa pasien yang tidak terdiagnosa diabetes dengan pasien yang terdiagnosa diabetes. Pengklasifikasian data diagnosa pasien terkait dengan diabetes dapat dilihat pada Gambar 1 berikut.



Gambar 1. Bobot Data Diagnosa Diabetes (a) Asli (b) Seimbang

Berdasarkan Gambar 1(a) di atas dapat dilihat bahwa perbandingan data pasien yang terdiagnosa diabetes dan tidak terdiagnosa diabetes tidak seimbang. Suatu data dikatakan tidak seimbang apabila proporsi data minoritas kurang dari 35% (Andiriani dan Susilaningrum, 2023). Oleh karena itu, data yang digunakan pada penelitian ini merupakan data yang tidak seimbang, sehingga diperlukan suatu penanganan untuk mengatasi ketidakseimbangan data tersebut. Salah satu penanganan yang dapat dilakukan yaitu dengan melakukan *random oversample* untuk data minoritas. Selanjutnya pada Gambar 1(b) dapat dilihat bahwa data yang digunakan telah memenuhi syarat keseimbangan data. Selanjutnya, data yang telah seimbang dapat digunakan dalam tahap selanjutnya.

Peubah bebas yang digunakan pada penelitian ini memiliki ukuran satuan yang berbeda. Peubah bebas *Pregnancies* menunjukkan jumlah kehamilan pada individu, *Glucose* menunjukkan kadar glukosa dalam darah individu, Insulin menunjukkan kadar Insulin dalam darah individu, *Body Mass Index* (BMI) menunjukkan Indeks Masa Tubuh individu, selanjutnya *Age* menunjukkan usia individu. Berdasarkan makna peubah bebas tersebut, terlihat bahwa skala nilai setiap peubah bebas berbeda. Oleh karena itu, diperlukan adanya penyelarasan skala data. Normalisasi *Min-Max* merupakan salah satu metode yang dapat digunakan untuk menyelaraskan skala data.



Gambar 2. *Boxplot* (a) *Pregnancies* (b) *Glucose* (c) Insulin (d) BMI (e) *Age*

Pada setiap data peubah bebas yang digunakan pada penelitian ini terdapat *outlier*. Hal tersebut dapat terlihat pada Gambar 2 yang menunjukkan *outlier* pada data peubah bebas. Namun, tidak semua *outlier* harus ditangani dengan cara dibuang. Pada dunia kesehatan, penanganan *outlier* perlu diperhatikan dengan hati-hati. *Outlier* pada data kesehatan bisa saja memberikan informasi klinis yang signifikan. Oleh karena itu, pada penelitian ini tidak dilakukan penanganan data *outlier*.

Kinerja Metode KNN, *Random Forest*, dan *Decision Tree* Dalam Memprediksi Diabetes

Berdasarkan hasil eksplorasi data diketahui bahwa peubah bebas yang digunakan pada penelitian ini memiliki skala yang berbeda. Pada penelitian ini penyelerasan data dilakukan dengan normalisasi Min-Max. Selain itu, pada penelitian ini dilakukan penanganan ketidakseimbangan data. Oleh karena itu, data yang digunakan pada penelitian ini adalah data yang telah seimbang. Perbandingan kinerja metode KNN, *Random Forest*, dan *Decision Tree* sebelum dilakukannya normalisasi dapat dilihat pada Tabel 5 berikut.

Tabel 2. Perbandingan Kinerja Metode Sebelum Normalisasi *Min-Max*

Metode	Akurasi	Waktu (detik)
KNN	0.740741	1.573881
Random Forest	0.780104	4.725485
Decision Tree	0.793548	3.483290

Berdasarkan Tabel 2 di atas, dapat dilihat akurasi kinerja setiap model beserta dengan waktu yang diperlukan dalam membangun masing-masing model di Julia pada data yang belum dilakukan normalisasi. *Decision Tree* memiliki nilai akurasi yang lebih tinggi dibandingkan metode-metode lainnya. Selanjutnya KNN memiliki waktu komputasi yang lebih cepat dibandingkan metode-metode lainnya. Pemilihan metode disesuaikan dengan tujuan penelitian yang akan dilakukan. Perbandingan kinerja metode pada data yang telah dilakukan normalisasi Min-Max dapat dilihat pada Tabel 6 berikut.

Tabel 3. Perbandingan Kinerja Metode Setelah Normalisasi *Min-Max*

Metode	Akurasi	Waktu (detik)
KNN	0.775281	1.528750
Random Forest	0.754601	4.628519
Decision Tree	0.737143	3.268580

Berdasarkan Tabel 3 dapat dilihat bahwa normalisasi Min-Max tidak memberikan peningkatan akurasi pada metode *Decision Tree* dan *Random Forest*. Namun dari sisi waktu komputasi, metode *Decision Tree* dan KNN memiliki waktu komputasi yang lebih kecil setelah dilakukan normalisasi Min-Max. Pada metode KNN, penanganan normalisasi Min-Max memberikan akurasi yang lebih tinggi dibandingkan sebelum dilakukan normalisasi. Selain itu, waktu komputasi metode KNN pada data yang telah dilakukan normalisasi Min-Max lebih kecil dibandingkan sebelum dilakukan normalisasi Min-Max.

PENUTUP

Simpulan

Metode *K-Nearest Neighbor* (KNN), *Random Forest*, dan *Decision Tree* dapat digunakan untuk memprediksi diabetes. Data peubah bebas yang digunakan pada penelitian ini memiliki skala ukuran yang berbeda, sehingga pada penelitian ini dilakukan normalisasi Min-Max. Hasil penelitian pada data yang belum dinormalisasi dengan Min-Max menunjukkan bahwa metode *Decision Tree* mem-iliki nilai akurasi yang lebih tinggi dibandingkan *K-Nearest Neighbor* (KNN) dan *Random Forest*. Sedangkan berdasarkan waktu komputasi, metode *K-Nearest Neighbor* (KNN) memiliki waktu komputasi yang lebih kecil dibandingkan *Decision Tree* dan *Random Forest*.

Pada data yang telah dilakukan normalisasi Min-Max, metode KNN mengalami peningkatan akurasi dan memiliki nilai akurasi yang lebih baik dibandingkan dengan metode *Decision Tree* dan *Random Forest*. Hasil metode *Decision Tree* dan *Random Forest* memiliki nilai akurasi yang berbanding terbalik dengan KNN pada data setelah normalisasi Min-Max. Sedangkan berdasarkan waktu komputasinya, ketiga metode menunjukkan waktu komputasi yang lebih cepat setelah dilakukan normalisasi Min-Max pada data peubah bebas yang digunakan.

Saran

Pada penelitian ini hanya digunakan lima peubah bebas, penelitian selanjutnya disarankan untuk menambahkan peubah bebas lainnya yang berkaitan dengan penyebab penyakit diabetes seperti jenis kelamin, kebiasaan konsumsi alkohol, dan lainnya. Selain itu, terdapat banyak metode lainnya yang dapat dilakukan untuk melakukan prediksi diagnosa diabates.

DAFTAR PUSTAKA

Alifah, R. N., Najib, M. K., Nurdiati, S., Sari, A. P., Herlambang, K., Ginting, D. T. P. B., & Sya'adah, S. N. (2024). Perbandingan Metode Tree Based Classification untuk Masalah Klasifikasi Data Body Mass Index. *Indonesian Journal of Mathematics and Natural Sciences*, 47(1), 49-65.

Andiriani, C. M. F., Susilaningrum, D. (2023). Klasifikasi Waiting Time for Pilot di Pelabuhan Tanjung Perak Menggunakan Metode Regresi Logistik - Synthetic Minority Oversampling Technique (SMOTE), *Jurnal Sains dan Seni ITS.* 12(1), 111-118.

Aprilliandhika, W., Abdulloh, F. F. (2024). Comparison Of K-Nearest Neighbor and Support Vector Machine Algorithm Optimization With Grid Search CV On Stroke Prediction*, Jurnal Teknik Informatika*, 5(4), 991-1000.

Aziz, M. I., Fanani, A. Z., Affandy. (2023). Analisis Metode Ensemble Pada Klasifikasi Penyakit Jantung Berbasis Decision Tree*, Jurnal Media Informatika Budidarma*, 7(1), 1-12.

Chairunisa, R., Adiwijaya, Astuti, W. (2020). Perbandingan CART dan *Random Forest* untuk Deteksi Kanker berbasis Klasifikasi Data *Microarray*. *Jurnal Rekayasa Sistem dan Teknologi Informasi.* 4(5), 805-812.

Gharehbaghi, A. (2023). *Deep Learning in Time Series Analysis*, Ed ke-1. Boca Raton (USA): CRC Press.

Handayani, P., Fauzan, A. C., Harliana. (2024). Machine Learning Klasifikasi Status Gizi Balita Menggunakan Algoritma Random Forest*, KLIK: Kajian Ilmiah Informatika dan Komputer,* 4(6), 3064-3072.

Hawari, I.F., Najib, M.K., Nurdiati, S., Marpaung, Y.F.Y., Kusumawati, N., Nurfadila, M., Sijabat, K.R. and Hernawan, B.F., (2024). Pengaruh Teknik Oversampling Pada Algoritma Machine Learning Dalam Klasifikasi Body Mass Index (BMI). *Jurnal Riset dan Aplikasi Matematika (JRAM)*, 8(1), pp.51-68.

Kozak J. (2019). *Decision Tree and Ensemble Learning Based on Ant Colony Optimization*, Cham(CH): Springer Nature Switzerland AG.

Maskuri, M. N., Harliana, Sukerti, K., Bhakti, R. M. H. (2022). Penerapan Algoritma K-Nearest Neighbor (KNN) untuk Memprediksi Penyakit Stroke, *Jurnal Ilmiah Intech: Information Technology Journal of UMUS,* 4(1), 130-140.

Pratama, Y., Prayitno, A., Nazrian, D., Rizki, Y., Rasywir, E. (2022). Klasifikasi Penyakit Gagal Jantung Menggunakan Algoritma K-Nearest Neighbor, *Bulletin of Computer Science Research.* 3(1), 52-56.

Rahman, I. F., Shestia, W. A., Rama, S. D., Azzahra, S., Dermawan, A. A. (2024). Klasifikasi Diagnosa Pasien Di Klinik Sri Dengan Metode Decision Tree, *Jurnal Teknik Ibnu Sina,* 9(1), 74-82.

Setianto, Y. A., Kusrini, K., Henderi, H. (2018). Penerapan Algoritma K-Nearest Neighbour Dalam Menentukan Pembinaan Koperasi Kabupatem Kotawaringin Timur, *Citec Journal,* 5(3), 232-241.

Solahuddin. M., Purnamasari, A. I., Dikananda, A. R. (2023). Klasifikasi Kualitas Berita Pada Majalah Menggunakan Metode Decision Tree*, Jurnal Teknologi Ilmu Komputer.* 1(2), 48-54.

Telaumbanua, F. D., Hulu, P., Nadeak, T. Z., Lumbantong, R. R., Dharma, A. (2019). Penggunaan Machine Learning Di Bidang Kesehatan, *Jurnal Penelitian Teknik Informatika,* 2(2), 391-399.

Wardhana, R. G., Wang, G., Sibuea, F. (2023). Penerapan Machine Learning Dalam Prediksi Tingkat Kasus Penyakit Di Indonesia, *Journal of Information System Management,* 5(1), 40-45.

[WHO] World Health Organization. (2023). Diabetes. Tersedia pada: https://www.who.int/news-room/fact-sheets/detail/diabetes