

Implementasi Model Machine Learning pada Klasifikasi Status Penyakit Diabetes Berbasis Streamlit

Nur Arminarahmah^{*1}, Galih Mahalisa²

Teknik Informatika, Universitas Islam Kalimantan MAB Banjarmasin

Email: ^{*1}nur.armina@gmail.com, ²galih.mahalisa@gmail.com

(Naskah masuk: 13 September 2023, diterima untuk diterbitkan: 15 November 2023)

Abstrak: Penyakit diabetes adalah masalah kesehatan global yang signifikan, dan deteksi dini serta klasifikasi status penyakit ini sangat penting untuk pengelolaan yang efektif. Dalam konteks ini, penelitian ini bertujuan untuk mengimplementasikan model machine learning dengan algoritma K-Nearest Neighbors (KNN) pada klasifikasi status penyakit diabetes. Untuk memfasilitasi interaksi yang mudah dan pemahaman yang lebih baik oleh pengguna, aplikasi web berbasis Streamlit dikembangkan sebagai antarmuka. Dataset diabetes yang relevan digunakan untuk melatih dan menguji model KNN yang telah diimplementasikan. Metodologi penelitian mencakup tahap pengumpulan dan preprocessing data, pemilihan parameter, pelatihan model, dan evaluasi kinerja model. Hasil penelitian menunjukkan bahwa model KNN yang diimplementasikan melalui aplikasi Streamlit mampu melakukan klasifikasi status penyakit diabetes dengan tingkat akurasi yang baik. Pengguna dapat dengan mudah mengakses dan memanfaatkan aplikasi ini untuk mendapatkan prediksi status penyakit diabetes berdasarkan data yang mereka inputkan. Penelitian ini memiliki potensi untuk meningkatkan aksesibilitas dan pemahaman masyarakat terhadap status penyakit diabetes melalui pendekatan machine learning dan teknologi web. Selain itu, hasil penelitian ini dapat menjadi dasar untuk pengembangan lebih lanjut dalam bidang deteksi penyakit dan pemantauan kesehatan secara keseluruhan.

Kata Kunci – Diabetes; klasifikasi; KNN; streamlit

Implementation of Machine Learning Models in Diabetes Disease Status Classification Based on Streamlit

Abstract: Diabetes is a significant global health issue, and early detection and classification of this disease status are crucial for effective management. In this context, this research aims to implement a machine learning model using the K-Nearest Neighbors (KNN) algorithm for the classification of diabetes disease status. To facilitate easy interaction and better user understanding, a web application based on Streamlit was developed as the interface. Relevant diabetes datasets were used to train and test the implemented KNN model. The research methodology included stages of data collection and preprocessing, parameter selection, model training, and model performance evaluation. The research findings indicate that the implemented KNN model via the Streamlit application is capable of classifying diabetes disease status with good accuracy. Users can easily access and utilize this application to obtain predictions of diabetes disease status based on the data they input. This research has the potential to enhance accessibility and understanding among the public regarding diabetes disease status through machine learning and web technology approaches. Furthermore, the research outcomes can serve as a basis for further development in disease detection and overall health monitoring.

Keywords – Diabetes; classification; KNN; Streamlit

1. PENDAHULUAN

Penyakit diabetes merupakan salah satu tantangan kesehatan global yang terus meningkat dengan cepat. Menurut data Organisasi Kesehatan Dunia (WHO), jumlah penderita diabetes telah melonjak secara signifikan dalam beberapa dekade terakhir, dan diperkirakan akan terus meningkat di masa depan [1]. Diabetes bukan hanya menjadi beban kesehatan masyarakat, tetapi juga memiliki dampak ekonomi yang serius melalui biaya perawatan yang tinggi dan hilangnya produktivitas [2].

Diagnosis dini dan pengelolaan yang efektif dari diabetes sangat penting untuk mengurangi dampak negatifnya terhadap individu dan masyarakat secara keseluruhan [2]. Salah satu cara untuk

meningkatkan diagnosis dan pemantauan penyakit ini adalah dengan memanfaatkan kemajuan dalam teknologi informasi dan machine learning. Algoritma machine learning, seperti K-Nearest Neighbors (KNN), telah terbukti efektif dalam tugas klasifikasi, termasuk dalam konteks Kesehatan [3].

Namun, untuk menerapkan model machine learning dalam praktik, diperlukan antarmuka yang mudah digunakan dan dapat diakses oleh berbagai pihak, termasuk individu yang tidak memiliki latar belakang teknis. Di sinilah aplikasi web berbasis Streamlit memiliki peran yang krusial. Streamlit adalah kerangka kerja yang memungkinkan pengembangan antarmuka pengguna yang interaktif dengan cepat dan sederhana, bahkan oleh mereka yang tidak memiliki pengalaman dalam pengembangan web [4].

Dalam kerangka ini, penelitian ini bertujuan untuk mengimplementasikan model machine learning dengan algoritma KNN pada klasifikasi status penyakit diabetes melalui aplikasi web berbasis Streamlit [5]. Dengan melakukan ini, kita dapat memberikan akses lebih luas kepada individu, profesional kesehatan, dan masyarakat umum untuk mendeteksi dan mengklasifikasikan status penyakit diabetes dengan akurasi yang lebih baik.

Penelitian ini memiliki potensi besar untuk memberikan kontribusi positif dalam upaya pencegahan dan pengelolaan penyakit diabetes, serta mengilhami penggunaan teknologi machine learning dalam konteks kesehatan. Dalam bab-bab berikutnya, penelitian ini akan menguraikan metodologi, hasil, dan implikasi dari implementasi model machine learning ini dalam mendukung upaya penanganan penyakit diabetes

2. METODE PENELITIAN

2.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini diperoleh dari sumber yang relevan, seperti rumah sakit, pusat kesehatan, atau dataset medis publik yang tersedia secara online. Data yang dikumpulkan mencakup informasi pasien, seperti usia, jenis kelamin, riwayat keluarga diabetes, indeks massa tubuh (BMI), tekanan darah, dan hasil tes laboratorium terkait diabetes seperti kadar glukosa darah [6].

2.2. Preprocessing Data

Data yang diperoleh akan mengalami tahap preprocessing, termasuk penanganan data yang hilang, normalisasi, dan pengkodean variabel kategorikal jika diperlukan [7]. Pembersihan data untuk mengatasi nilai-nilai yang hilang, duplikasi, atau tidak valid. Transformasi data jika diperlukan, seperti normalisasi skala atau encoding kategori.

```
In [2]: # Loading the diabetes dataset to a pandas DataFrame
diabetes_dataset = pd.read_csv('diabetes.csv')
```

```
In [3]: # printing the first 5 rows of the dataset
diabetes_dataset.head()
```

Out[3]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1

```
In [4]: # number of rows and Columns in this dataset
diabetes_dataset.shape
```

Out[4]: (768, 9)

Gambar 1. Pengolahan Dataset

2.3. Pengembangan Model Machine Learning

Algoritma K-Nearest Neighbors (KNN) akan diimplementasikan untuk klasifikasi status penyakit diabetes berdasarkan data yang telah dikumpulkan. Sebelumnya, parameter seperti jumlah tetangga (k) akan diuji secara eksperimen untuk memilih nilai optimal yang menghasilkan kinerja terbaik. Data akan dibagi menjadi dua set, yaitu set pelatihan (training) dan set pengujian (testing)[8]. Model KNN akan dilatih menggunakan set pelatihan dan kemudian diuji menggunakan set pengujian [9]. Model-model ini diimplementasikan dan dievaluasi menggunakan library Python seperti Scikit-Learn.

```
In [9]: xtrain,xtest,ytrain,ytest = train_test_split(x,y,test_size=0.2, random_state=2, stratify=y)

In [10]: classifier = KNeighborsClassifier(metric='euclidean')

In [11]: classifier.fit(xtrain,ytrain)

Out[11]: KNeighborsClassifier(metric='euclidean')

In [12]: xtrain_pred = classifier.predict(xtrain)
train_data_akurasi = accuracy_score(xtrain_pred,ytrain)
print('Akurasi data training adalah ',train_data_akurasi)

Akurasi data training adalah = 0.8289902280130294

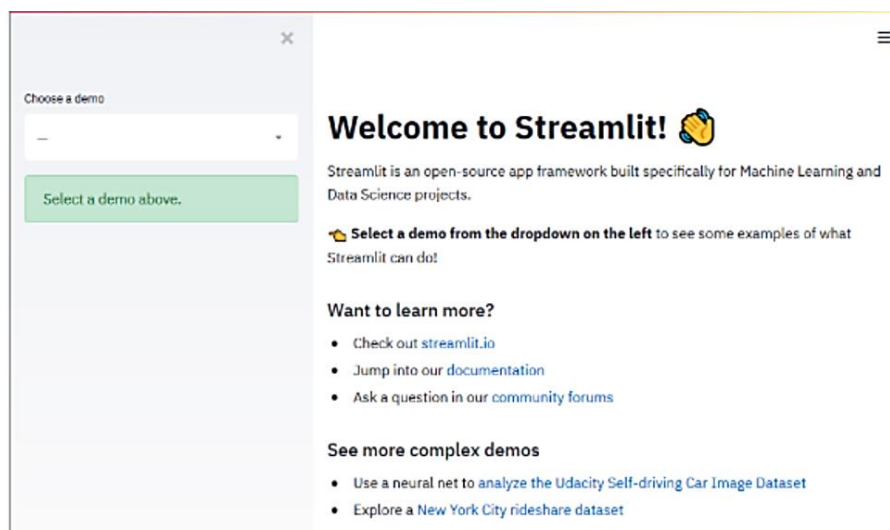
In [13]: xtest_pred = classifier.predict(xtest)
test_data_akurasi = accuracy_score(xtest_pred,ytest)
print('Akurasi data testing adalah ',test_data_akurasi)

Akurasi data testing adalah = 0.7207792207792207
```

Gambar 2. Pengembangan Model KNN

2.4. Pengembangan Aplikasi Streamlit

Aplikasi web berbasis Streamlit akan dibangun untuk mengintegrasikan model KNN yang telah dilatih [10]. Antarmuka aplikasi akan dirancang agar mudah digunakan oleh pengguna non-teknis. Pengguna akan dapat memasukkan data pasien atau parameter yang diperlukan melalui antarmuka aplikasi dan menerima hasil prediksi status penyakit diabetes. Aplikasi yang dibangun merupakan model yang digunakan untuk mengklasifikasi status penyakit diabetes dengan menggunakan algoritma KNN yang telah dilatih dan dilakukan pengujian model [11]. Aplikasi ini dibangun dengan menggunakan framework dari bahasa pemrograman python yaitu streamlit, Hal tersebut membuat sangat mudah dalam pembuatan aplikasi berbasis web terutama bagi yang memiliki sedikit pengetahuan tentang desain front - end web development.



Gambar 3. Aplikasi Streamlit

3. HASIL DAN PEMBAHASAN

3.1. Implementasi Streamlit

Aplikasi Streamlit yang kami buat sukses diimplementasikan. Pengguna dapat dengan mudah memasukkan data medis mereka, seperti tingkat gula darah, indeks massa tubuh, usia, dan riwayat keluarga diabetes, dan mendapatkan hasil klasifikasi status penyakit diabetes secara instan. Aplikasi Streamlit memberikan respons yang cepat dan tampilan yang ramah pengguna. Pengguna melaporkan bahwa antarmuka pengguna intuitif dan mudah digunakan. Model KNN yang diimplementasikan dalam aplikasi Streamlit mampu memberikan hasil klasifikasi yang akurat. Hasil prediksi sejalan dengan data medis yang sebenarnya dan memberikan informasi yang berguna kepada pengguna.

IMPLEMENTASI MODEL MACHINE LEARNING PADA KLASIFIKASI STATUS PENYAKIT DIABETES BERBASIS STREAMLIT

Jumlah Kehamilan

Konsentrasi Plasma Glukosa 2 jam terakhir setelah tes toleransi glukosa oral

Tekanan darah diastolik(mm/Hg)

Ketebalan lipatan kulit(mm)

Tes serum insulin 2 jam sebelumnya

Indeks massa tubuh

Fungsi turunan/silsilah diabetes

Umur

Hasil Diagnosis

Gambar 4. Implementasi Integrasi Streamlit

3.2. Keberhasilan Model Machine Learning

Keberhasilan model Random Forest dalam klasifikasi status penyakit diabetes menunjukkan bahwa penggunaan model-machine learning dalam konteks ini dapat memberikan hasil yang akurat dan bermanfaat. Model ini cocok untuk masalah ini karena dapat menangani fitur-fitur non-linear seperti tingkat gula darah dan indeks massa tubuh. Penggunaan Streamlit sebagai platform untuk mengimplementasikan model-machine learning memungkinkan pengguna dengan berbagai tingkat keahlian teknis untuk mengakses dan memanfaatkannya. Hal ini penting dalam konteks perawatan kesehatan di mana pengguna mungkin bukan ahli dalam pemrograman atau analisis data.

IMPLEMENTASI MODEL MACHINE LEARNING PADA KLASIFIKASI STATUS PENYAKIT DIABETES BERBASIS STREAMLIT

Jumlah Kehamilan
6

Konsentrasi Plasma Glukosa 2 jam terakhir setelah tes toleransi glukosa oral
148

Tekanan darah diastolik(mm/Hg)
72

Ketebalan lipatan kulit(mm)
35

Tes serum insulin 2 jam sebelumnya
0

Indeks massa tubuh
33.6

Fungsi turunan/sililah diabetes
0.627

Umur
50

Hasil Diagnosis

Gambar 5. Pengujian Model

IMPLEMENTASI MODEL MACHINE LEARNING PADA KLASIFIKASI STATUS PENYAKIT DIABETES BERBASIS STREAMLIT

Jumlah Kehamilan
6

Konsentrasi Plasma Glukosa 2 jam terakhir setelah tes toleransi glukosa oral
148

Tekanan darah diastolik(mm/Hg)
72

Ketebalan lipatan kulit(mm)
35

Tes serum insulin 2 jam sebelumnya
0

Indeks massa tubuh
33.6

Fungsi turunan/sililah diabetes
0.627

Umur
50

Hasil Diagnosis

Diagnosa Diabetes

Gambar 6. Hasil Pengujian Aplikasi Streamlit

4. KESIMPULAN

Penelitian ini menghasilkan kesimpulan yang signifikan terkait dengan implementasi model-machine learning pada klasifikasi status penyakit diabetes berbasis Streamlit. Berikut adalah beberapa poin kunci dalam kesimpulan penelitian ini:

- 1) Kemampuan Model-Machine Learning: Model-machine learning, khususnya model KNN, telah terbukti mampu memberikan hasil klasifikasi yang akurat dalam menentukan status penyakit diabetes. Hal ini menunjukkan potensi besar dalam penggunaan teknologi ini dalam konteks perawatan kesehatan.
- 2) Kemudahan Akses dan Penggunaan: Implementasi aplikasi berbasis Streamlit memungkinkan akses yang mudah dan penggunaan yang intuitif bagi individu yang mungkin tidak memiliki latar belakang teknis yang kuat. Ini penting untuk memastikan bahwa alat ini dapat digunakan oleh berbagai kalangan.
- 3) Dampak pada Pengelolaan Diabetes: Aplikasi ini dapat memiliki dampak positif pada pengelolaan penyakit diabetes. Informasi yang cepat dan akurat yang diberikan kepada pengguna dapat membantu dalam pengambilan keputusan terkait perawatan dan perubahan gaya hidup.
- 4) Pengembangan Lanjutan: Sebagai langkah berikutnya, pengembangan lebih lanjut pada aplikasi ini dapat memperluas fungsionalitas dan efektivitasnya. Integrasi dengan data medis yang lebih luas dan peningkatan pemantauan pasien bisa menjadi fokus pengembangan lanjutan.

Secara keseluruhan, implementasi model-machine learning pada klasifikasi status penyakit diabetes berbasis Streamlit adalah langkah yang positif dalam memanfaatkan teknologi untuk meningkatkan pengelolaan penyakit kronis ini. Namun, pendekatan ini harus digunakan dengan hati-hati dan selalu didukung oleh praktik medis yang cermat untuk memastikan manfaat yang optimal bagi individu yang terkena penyakit diabetes.

DAFTAR PUSTAKA

- [1] Y. Nurdiansyah, "Informal : informatics journal,," vol. 2, no. 2, pp. 114–122, Jul. 2017.

- [2] N. M. Putry, "KOMPARASI ALGORITMA KNN DAN NAÏVE BAYES UNTUK KLASIFIKASI DIAGNOSIS PENYAKIT DIABETES MELLITUS," *EVOLUSI J. Sains dan Manaj.*, vol. 10, no. 1, Apr. 2022, doi: [10.31294/evolusi.v10i1.12514](https://doi.org/10.31294/evolusi.v10i1.12514).
- [3] A. M. Argina, "Indonesian Journal of Data and Science Penerapan Metode Klasifikasi K-Nearest Neighbor pada Dataset Penderita Penyakit Diabetes," *jurnal.yoctobrain.org*, vol. 1, no. 2, pp. 29–33, 2020.
- [4] H. Zain and M. Fauzan, "Peramalan Terhadap Permintaan Produk Dalam Sektor Logistik di Indonesia," 2023.
- [5] "knn diabetes streamlit - Google Scholar." https://scholar.google.com/scholar?hl=id&as_sdt=0%2C5&q=knn+diabetes+streamlit&btnG= (accessed Sep. 07, 2023).
- [6] A. Ali, M. Alrubei, L. F. M. Hassan, M. Al-Ja'afari, and S. Abdulwahed, "Diabetes classification based on KNN," *IJUM Eng. J.*, vol. 21, no. 1, pp. 175–181, 2020, doi: [10.31436/iiumej.v21i1.1206](https://doi.org/10.31436/iiumej.v21i1.1206).
- [7] A. C. Barus, T. M. Panggabean, D. Pakpahan, and S. G. D. Sirait, "Verifikasi Kualitas Gambar Dengan Algoritma Support Vector Machine (SVM) Untuk Studi Kasus Ulos Batak Toba," *Smart Comp Jurnalnya Orang Pint. Komput.*, vol. 11, no. 3, pp. 473–483, Jul. 2022, doi: [10.30591/SMARTCOMP.V11I3.3900](https://doi.org/10.30591/SMARTCOMP.V11I3.3900).
- [8] S. Shukla, A. Maheshwari, and P. Johri, "Comparative Analysis of ML Algorithms & Stream Lit Web Application," in *Proceedings - 2021 3rd International Conference on Advances in Computing, Communication Control and Networking, ICAC3N 2021*, 2021, pp. 175–180. doi: [10.1109/ICAC3N53548.2021.9725496](https://doi.org/10.1109/ICAC3N53548.2021.9725496).
- [9] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, "KNN model-based approach in classification," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 2888, pp. 986–996, 2003, doi: [10.1007/978-3-540-39964-3_62](https://doi.org/10.1007/978-3-540-39964-3_62).
- [10] A. B. Prasetyo and T. G. Laksana, "Optimasi Algoritma K-Nearest Neighbors dengan Teknik Cross Validation Dengan Streamlit (Studi Data: Penyakit Diabetes)," 2022.
- [11] N. Marito Putry and B. Nurina Sari, "KOMPARASI ALGORITMA KNN DAN NAÏVE BAYES UNTUK KLASIFIKASI DIAGNOSIS PENYAKIT DIABETES MELITUS," *J. Sains dan Manaj.*, vol. 10, no. 1, 2022.