

Implementasi Data Mining dengan Algoritma K-Nearest Neighbors untuk Memprediksi Risiko Diabetes Menggunakan Chatbot Telegram

Diterima:

10 Juni 2024

Revisi:

10 Juli 2024

Terbit:

1 Agustus 2024

^{1*}Muhammad Reza Muzaki, ²Intan Melinda Nurfajriana, ³Ferlita Putri Anugerah Ilahi, ⁴Aidina Ristyawan, ⁵Erna Daniati

¹⁻⁵Universitas Nusantara PGRI Kediri

¹mrreza545@gmail.com, ²intanmelindanurfajriana@gmail.com,

³lilypai804@gmail.com, ⁴adinaristi@unpkediri.ac.id,

⁵ernadaniati@unpkediri.ac.id

Abstrak— Diabetes merupakan penyakit kronis yang disebabkan karena tingginya kadar gula dalam darah. Menurut *International Diabetes Federation (IDF)* memperkirakan pada tahun 2045, angka penderita diabetes akan meningkat menjadi 11,3% pada keseluruhan total penduduk di Indonesia. Penyakit diabetes dengan cepat menyebar dalam tubuh seseorang dengan cepat, dan banyak masyarakat yang menyadari akan hal ini. Dengan integrasi kedalam *chatbot* telegram di harapkan bisa membantu untuk memprediksi risiko diabetes. Model data mining menggunakan algoritma KNN mendapatkan hasil akurasi sebesar 95%, recall sebesar 99%, dan presisi sebesar 89,1%

Kata Kunci—Data Mining; K-Nearest Neighbors; Diabetes; Chatbot

Abstract— *Diabetes is a chronic disease caused by high levels of sugar in the blood. According to the International Diabetes Federation (IDF), by 2045, the number of people with diabetes will increase to 11.3% of the total population in Indonesia. Diabetes quickly spreads in a person's body quickly, and many people are aware of this. With the integration into the telegram chatbot, it is hoped that it can help to predict the risk of diabetes. The data mining model using the KNN algorithm obtained accuracy results of 95%, recall of 99%, and precision of 89.1%.*

Keywords— *Data Mining; K-Nearest Neighbors; Diabetes; Chatbot*

This is an open access article under the CC BY-SA License.



Penulis Korespondensi:

Aidina Ristyawan,

Sistem Informasi,

Universitas Nusantara PGRI Kediri,

Email: adinaristi@unpkediri.ac.id,

ID Orcid: [<https://orcid.org/0009-0003-2712-1507>]

Handphone: 081232624460

I. PENDAHULUAN

Diabetes merupakan masalah kesehatan utama yang saat ini telah mencapai tingkat yang mengkhawatirkan. Pada tahun 2019, di seluruh dunia hampir ada setengah miliar orang (9,3% orang dewasa berusia 20-79 tahun) hidup dengan menderita penyakit diabetes. Telah diperkirakan jumlah orang usia 20-79 tahun yang hidup dengan diabetes akan meningkat sebesar 62% selama 10 tahun terakhir, dari 285 juta penderita pada tahun 2009 menjadi 463 juta penderita pada tahun ini. Saat ini, setengah (50,1%) penderita diabetes tidak mengetahui bahwa dirinya mengidap diabetes [1].

Diabetes adalah gangguan metabolisme energi, protein, dan lemak pada tubuh disebabkan oleh peningkatan resistensi insulin yang mengakibatkan terganggunya sintesis insulin [2]. Penyakit diabetes disebabkan oleh peningkatan gula darah (glukosa) atau kadar gula dalam darah. Pemantauan kadar gula dalam darah harus selalu dipantau dan dikelola dengan baik, apabila tidak dikelola dengan baik akan dapat menyebabkan diabetes, sehingga dapat menyebabkan komplikasi berbagai penyakit, seperti obesitas, stroke, stroke, jantung koroner, serta gangguan pada fungsi ginjal, saraf, dan penglihatan [3].

Pengecekan diabetes secara dini diperlukan agar dapat dengan cepat ditangani dan menurunkan risiko terjadinya komplikasi. Prediksi risiko diabetes dimaksudkan untuk membantu pengguna memahami risiko terkena penyakit diabetes atau untuk menentukan potensi terkena penyakit tersebut [4]. Pada prediksi pada penelitian ini, didasarkan pada algoritma K-Nearest Neighbor (KNN) yang mengumpulkan data dan menganalisisnya menggunakan beberapa parameter. Ini termasuk kondisi kompleks seperti jantung koroner, stroke, dan obesitas.

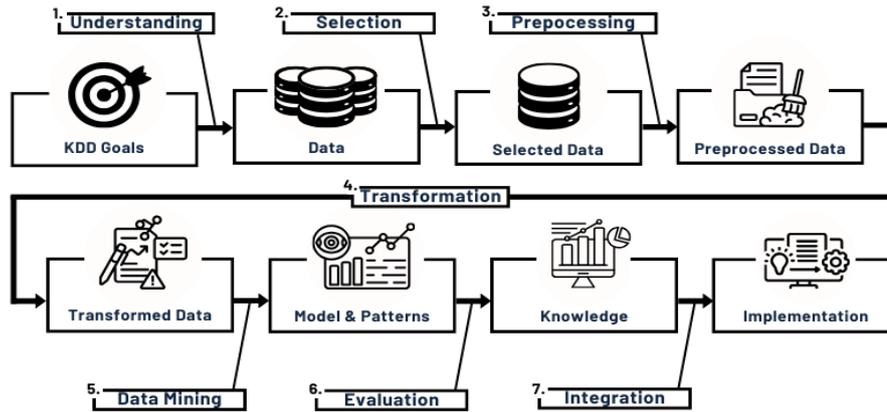
Algoritma K-Nearest Neighbors (KNN) merupakan metode klasifikasi sekelompok data berdasarkan pembelajaran data yang sudah terklasifikasikan sebelumnya. Metode KNN termasuk dalam *supervised learning*, yang dimana hasil *query instance* baru akan diklasifikasikan berdasarkan kedekatan jarak dari kategori yang ada dalam KNN [5]. Data Mining merupakan proses pencarian *single data* hingga menjadi *multiple data*, dengan otomatis sehingga mendapatkan sebuah model dari database.

Penggunaan algoritma KNN ini digunakan untuk pembuatan model dan integrasikan melalui *Chatbot* pada aplikasi Telegram. *Chatbot* merupakan sebuah program yang dibuat untuk menirukan cara komunikasi manusia yang interaktif melalui bentuk tulisan, suara, maupun visual [6]. *Chatbot* berguna untuk mempermudah dalam interaksi secara tidak langsung kepada pengguna, dengan menggunakan *Chatbot* pengambilan informasi atau data-data pengguna akan secara otomatis diolah dan diproses dengan model yang telah ditanamkan menjadi otak dari *bot*, sehingga dapat memunculkan prediksi risiko diabetes secara *real-time*.

Pada penelitian sebelumnya yang berjudul "Implementasi Data Mining untuk Memprediksi Penyakit Diabetes Menggunakan Algoritma Naives Bayes dan K-Nearest Neighbor" penelitian tersebut terkait dengan prediksi diabetes menggunakan 2 algoritma sebagai perbandingan dan menggunakan cross validation dengan 10 kali *fold* (lipatan). Dengan hasil yang diperoleh algoritma K-Nearest Neighbor yaitu 99% dan Naive Bayes yaitu 75%, model algoritma KNN lebih unggul dibandingkan dari algoritma Naive bayes [7].

II. METODE

Metode yang digunakan dalam penelitian ini adalah proses model data mining KDD. Proses KDD (Knowledge Discovery in Databases) merupakan metode untuk memperoleh pengetahuan dari database. Hasil pengetahuan tersebut dapat digunakan sebagai basis pengetahuan (*knowledge base*) untuk pengambilan keputusan atau dapat diintegrasikan dengan dengan sistem yang lain [8]. Proses model KDD dipilih dalam penelitian ini karena memiliki pendekatan yang sistematis untuk mengeksplorasi dan menganalisa data, sehingga dapat menghasilkan model yang dapat diandalkan dan mudah diintegrasikan ke dalam sistem yang lebih luas.



Gambar 1. Proses model data mining KDD

Pada gambar 1 menampilkan alur metode KDD yang digunakan untuk penelitian ini, dimulai dengan proses *understanding* (Menentukan *dataset* yang diperlukan untuk penelitian), *selection* (seleksi data yang akan digunakan), *preprocessing* (mempersiapkan data), *transformation* (mentransformasikan data), *data mining* (menggali informasi dari data), *evaluation* (mengevaluasi hasil penelitian), dan *integration* (pengintegrasian model data mining). Penjelasan lebih lanjut mengenai alur proses data mining KDD adalah sebagai berikut:

1. Understanding

Tujuan beserta target sasaran KDD ditentukan dari sudut pandang pengguna, hal ini digunakan untuk mengembangkan pemahaman tentang domain aplikasi dan pengetahuan sebelumnya [9]. pada penelitian ini kami menggunakan *dataset* diabetes dari situs web kaggle.com [10] yang memiliki sampel sejumlah 100.000 data. Pemrograman data menggunakan Bahasa pemrograman Python dengan aplikasi Jupyter Notebook dan menggunakan *library* seperti Pandas, Numpy, Sklearn, dan Matplotlib. *Data frame* dari *dataset* diabetes dapat dilihat pada gambar 2.

	gender	age	hypertension	heart_disease	smoking_history	bmi	HbA1c_level	blood_glucose_level	diabetes
0	Female	80.0	0	1	never	25.19	6.6	140	0
1	Female	54.0	0	0	No Info	27.32	6.6	80	0
2	Male	28.0	0	0	never	27.32	5.7	158	0
3	Female	36.0	0	0	current	23.45	5.0	155	0
4	Male	76.0	1	1	current	20.14	4.8	155	0
...
99995	Female	80.0	0	0	No Info	27.32	6.2	90	0
99996	Female	2.0	0	0	No Info	17.37	6.5	100	0
99997	Male	66.0	0	0	former	27.83	5.7	155	0
99998	Female	24.0	0	0	never	35.42	4.0	100	0
99999	Female	57.0	0	0	current	22.43	6.6	90	0

100000 rows x 9 columns]

Gambar 2. *Data frame* pada *dataset* diabetes

2. Data Selection

Tahap ini dilakukan dengan pemilihan data atau pemilihan atribut yang selanjutnya dilakukan *preprocessing* dan *data mining* [11]. Pada *dataset* yang telah ditentukan, terdapat 8 kolom atribut dan 1 kolom label yang dipilih untuk digunakan pada penelitian *data mining* ini, deskripsi kolom dapat dilihat pada tabel 1.

Tabel 1. Deskripsi Kolom

Kolom	Penjelasan
Gender	Jenis kelamin pasien
Age	Usia pasien pada saat diagnosis
hypertension	Riwayat tekanan darah tinggi
heart_disease	Riwayat Penyakit jantung
smoking_history	Riwayat merokok pasien
BMI	Indeks masa tubuh
HbA1c_level	kadar gula darah rata-rata dalam 2-3 bulan
blood_glucose_level	Kadar gula dalam darah
diabetes	Diagnosa diabetes

Pada tabel 1 menunjukkan deskripsi atau penejelasan dari 8 kolom atribut dan 1 kolom label yang ada pada *dataset* diabetes, peneliti memilih untuk mencantumkan keseluruhan kolom atribut untuk membantu mengetahui signifikansi setiap atribut tersebut terhadap kolom label (diabetes).

3. Preprocessing

Preprocessing data bertujuan untuk mengolah data *raw* (mentahan) menjadi data berkualitas dan siap untuk diolah pada tahap selanjutnya [12]. Beberapa tahapan dari *preprocessing* yang digunakan adalah penanganan data duplikat, *missing value*, dan *imbalance data* hal ini bertujuan untuk menyiapkan data supaya relevan, dan sebagai syarat awal dalam melakukan data mining. Penanganan data duplikat dilakukan dengan cara menampilkan banyaknya data yang sama (data redundan), dan melakukan penghapusan data tersebut.

```
[135]: print("____Jumlah Data Duplikat____")
print(df.duplicated().sum())

#Menghapus data duplicate
df = df.drop_duplicates()
print("____Menghapus Data Duplikat____")
print(df.duplicated().sum())

____Jumlah Data Duplikat____
3854
____Menghapus Data Duplikat____
0
```

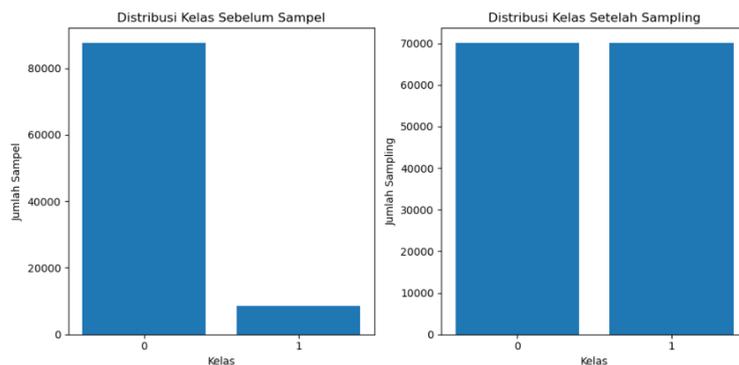
Gambar 3. Penanganan data duplikat

Pada gambar 3 terlihat di dalam *dataset* ini terdapat sebanyak 3.854 data duplikat, selanjutnya data tersebut dihapus sehingga sampel tersisa 96.146 data. Pengecekan *missing value* dilakukan dengan mencari apakah terdapat didalam data terdapat *value* (nilai) yang hilang atau kosong. *Missing value* dapat dilihat informasi data dari *dataset* diabetes, pada kolom 'Non-Null Count' akan menunjukkan jumlah sampel data yang tidak kosong (non-null) yang berarti tidak terdapat *missing value*, Informasi data dari dataset diabetes dapat dilihat pada gambar 4.

#	Column	Non-Null Count
0	gender	96146 non-null
1	age	96146 non-null
2	hypertension	96146 non-null
3	heart_disease	96146 non-null
4	smoking_history	96146 non-null
5	bmi	96146 non-null
6	HbA1c_level	96146 non-null
7	blood_glucose_level	96146 non-null
8	diabetes	96146 non-null

Gambar 4. Informasi data dari dataset diabetes

Selanjutnya untuk menangani *imbalance data* (data tidak seimbang) diperlukan pengecekan label data terlebih dahulu, untuk memastikan bahwa apakah persebaran *class* seimbang atau tidak. Pendistribusian kelas yang tidak seimbang akan menghasilkan model yang bias dengan kecenderungan memilih kelas mayoritas. Untuk mengatasi hal tersebut, akan dilakukan *oversampling* dan *undersampling* atau biasa disebut dengan *hybrid sampling*. *Oversampling* dilakukan menggunakan teknik SMOTE (Synthetic Minority Over-sampling Technique) yang akan menambahkan data pada kelas minoritas dengan menggunakan data sintetik, sehingga kelas minoritas akan dapat mengimbangi distribusi dari kelas mayoritas. Sedangkan *undersampling* menggunakan teknik ‘Tomek Links’ yang cara kerjanya menghapus sampel-sampel yang berdekatan antara kelas mayoritas dan kelas minoritas.



Gambar 5. Distribusi kelas sebelum dan sesudah *sampling*

Seperti yang ditunjukkan pada gambar 5, pendistribusian kelas sebelum dilakukan *sampling* sangat tidak seimbang dengan hasil 0 (negatif diabetes) sebesar 87.664 sampel data dan hasil 1 (positif diabetes) berjumlah 8.482 sampel data. Setelah dilakukan proses *sampling* distribusi kelas menjadi seimbang dengan hasil ‘0’ dan ‘1’ yang sama, yaitu sejumlah 70.000 sampel data.

4. Transformation

Tahap transformasi dilakukan untuk pemrosesan data skala, normalisasi, dan reduksi dimensi [13]. Transformasi pengubahan tipe data dari tipe non-numerik ke tipe data numerik harus dilakukan untuk dapat diproses pada saat *data mining*. Pada *dataset* diabetes ini terdapat 2 kolom atribut yang memiliki tipe data non-numerik, yaitu pada kolom ‘gender’ dan ‘smoking_history’ yang bertipe data *object*. Kolom ‘gender’ diubah menjadi data numerik yang mengelompokkan jenis kelamin laki-laki diibaratkan dengan angka 0 dan perempuan sebagai angka 1, sedangkan kolom ‘smoking_history’ dari pengelompokkan 6 kategori akan diambil nilai *value* nya seperti pada tabel 2.

Tabel 2. *Value* kategori pada kolom ‘smoking_history’

Kategori	Value
No Info	0
Current	1
Ever	2
Former	3
never	4
Not current	5

Setelah melakukan transformasi data, akan menghasilkan data yang bertipe numerik dan siap untuk tahap data mining, dataset diabetes setelah melalui tahap transformation dan *preprocessing* akan menghasilkan data seperti pada gambar 6.

gender	age	hypertension	heart_disease	smoking_history	bmi	HbA1c_level	blood_glucose_level	diabetes
0	80.0	0	1	4	25.19	6.6	140	0
0	54.0	0	0	0	27.32	6.6	80	0
1	28.0	0	0	4	27.32	5.7	158	0
0	36.0	0	0	1	23.45	5.0	155	0
1	76.0	1	1	1	20.14	4.8	155	0

Gambar 6. *Dataset* setelah dilakukan tahap *preprocessing*

Terlihat pada gambar 6, yang sebelumnya kolom ‘gender’ dan ‘smoking_history’ bertipe data ‘object’ (non-numerik) sekarang sudah bertipe data ‘int64’ (numerik) setelah dilakukannya proses *data transformation*. Keseluruhan data pada 9 kolom sudah bertipe numerik dan selanjutnya data sudah siap untuk melalui proses *data mining*.

5. *Data Mining*

Dalam proses Data Mining ditemukan pola-pola yang dapat digunakan untuk membangun model yang mampu memprediksi perilaku atau mengidentifikasi karakteristik umum dari objek yang sedang diteliti [14]. *Dataset* dibagi menjadi 2 variabel (X dan y) untuk memisahkan atribut dengan label, variabel ‘X’ berisi seluruh kolom atribut dan variabel ‘y’ berisi kolom label. Setelah itu akan ditentukan data latih dan data uji (X_train, X_test, y_train, y_test) dengan pembagian data uji 25% dan data latih sebesar 75%. Pelatihan dilakukan menggunakan algoritma KNN dengan parameter ‘n_neighbors = 7’ yang berarti model akan mempertimbangkan 7 tetangga terdekat dari titik data yang ingin diprediksi untuk menentukan kelasnya.

6. *Evaluation*

Tahap evaluasi dilakukan dengan tujuan untuk memastikan hasil analisis yang akurat dan dapat digunakan [15]. Dalam mengevaluasi algoritma KNN yang telah dijalankan pada proses *data mining*, akan ditunjukkan hasil-hasil dari pengukuran kinerja model tersebut seperti tingkat akurasi, presisi, dan *recall* untuk menganalisis tingkat efektivitas dari algoritma dalam suatu model yang dibuat. hasil perhitungan akan menegaskan apakah algoritma KNN akan menunjukkan tingkat efektivitas yang baik dalam memprediksi risiko diabetes, dan memberikan kepercayaan yang lebih kuat terhadap hasil prediksi untuk nantinya dilakukan proses pengintegrasian model.

7. *Integration*

Integrasi model prediksi risiko diabetes diimplementasikan kedalam aplikasi *chatbot* telegram untuk dapat dengan mudah dijangkau oleh khalayak umum, dan dapat melayani pengguna kapan saja karena chatbot tidak dibatasi oleh waktu [16]. *Chatbot* dapat dijalankan secara lokal menggunakan Python interpreter atau bisa dengan menggunakan server dari (phytonanywhere.com) sebagai tempat pemrosesan model prediksi yang telah dibuat.

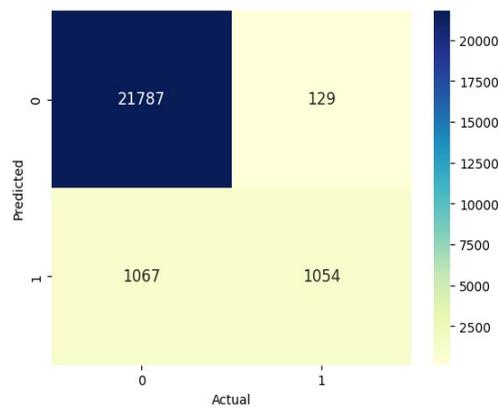
III. HASIL DAN PEMBAHASAN

Dari pemrosesan data mining pada dataset diabetes menggunakan *treatment data* seperti penanganan *missing value*, *label encoding*, *hybrid sampling*, dan *cross-validation* mendapatkan hasil seperti pada gambar 7.

	precision	recall	f1-score	support
0	0.95	0.99	0.97	21916
1	0.89	0.50	0.64	2121
accuracy			0.95	24037
macro avg	0.92	0.75	0.81	24037
weighted avg	0.95	0.95	0.94	24037

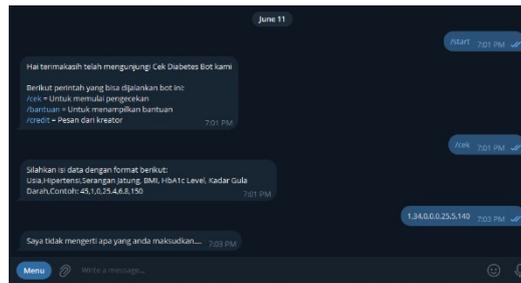
Gambar 7. Hasil *report* model

Hasil model KNN diimplementasikan dan menunjukkan performa yang baik dengan *recall* sebesar 75%, Precision sebesar 92%, dan Accuracy sebesar 95%. Dengan hasil grafik *confusion matrix* seperti pada gambar 8.



Gambar 8. Confusion matriks

Prototype Chatbot Telegram yang menggunakan jupyter notebook dengan menggunakan Bahasa pemrograman python di serialisasi dengan menggunakan *library* ‘pickle’ dan kemudian dijalankan kedalam server ‘phytonanywhere.com’. Tampilan *prototype chatbot* telegram bisa dilihat pada gambar 9.



Gambar 9. Tampilan telegram chatbot (Cek Diabetes Bot)

IV. KESIMPULAN

Data mining dari dataset diabetes dengan menggunakan algoritma KNN menunjukkan performa yang baik, dengan *recall* sebesar 99%, *Precision* sebesar 89,1%, dan *Accuracy* sebesar 95%. Hasil dari evaluasi menunjukkan bahwa algoritma KNN efektif dalam memprediksi risiko diabetes. Tingkat akurasi yang tinggi memberikan kepercayaan lebih besar terhadap hasil prediksi model, sehingga dapat digunakan untuk mendukung keputusan klinis dan strategi tindak lanjut yang lebih tepat bagi pasien. Dengan integrasi ke dalam chatbot telegram diharapkan dapat membantu dalam memprediksi risiko diabetes.

DAFTAR PUSTAKA

- [1] P. Saeedi et al., "Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition," *Diabetes Res Clin Pract*, vol. 157, p. 107843, Nov. 2019, doi: 10.1016/j.diabres.2019.107843.
- [2] C. Debora Mait, J. Armando Watuseke, P. David Gibrael Saerang, S. Reynaldo Joshua, and U. Sam Ratulangi, "Sistem Pendukung Keputusan Menggunakan Fuzzy Logic Tahani Untuk Penentuan Golongan Obat Sesuai Dengan," *Jurnal Media Infotama*, vol. 18, no. 2, pp. 344–353, 2022, doi: <https://doi.org/10.37676/jmi.v18i2.2936>.
- [3] A. M. Argina, "Penerapan Metode Klasifikasi K-Nearest Neighbor pada Dataset Penderita Penyakit Diabetes," *Indonesian Journal of Data and Science*, vol. 1, no. 2, pp. 29–33, 2020, doi: 10.33096/ijodas.v1i2.11.
- [4] Y. F. Wijaya and A. Triayudi, "Perbandingan Algoritma Klasifikasi Data Mining Pada Prediksi Penyakit Diabetes," *Journal of Computer System and Informatics (JoSYC)*, vol. 5, no. 1, pp. 165–174, 2023, doi: 10.47065/josyc.v5i1.4614.
- [5] D. Cahyanti, A. Rahmayani, and S. A. Husniar, "Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara," *Indonesian Journal of Data and Science*, vol. 1, no. 2, pp. 39–43, Jul. 2020, doi: 10.33096/IJODAS.VII2.13.
- [6] D. S. Hormansyah and Y. P. Utama, "APLIKASI CHATBOT BERBASIS WEB PADA SISTEM INFORMASI LAYANAN PUBLIK KESEHATAN DI MALANG DENGAN MENGGUNAKAN METODE TF-IDF," *Jurnal Informatika Polinema*, vol. 4, no. 3, pp. 224–228, May 2018, doi: 10.33795/JIP.V4I3.211.
- [7] F. N. Ikhromr, I. Sugiyarto, U. Faddillah, and B. Sudarsono, "Implementasi Data Mining Untuk Memprediksi Penyakit Diabetes Menggunakan Algoritma Naives Bayes dan K-Nearest Neighbor," *INTECOMS: Journal of Information Technology and Computer Science*, vol. 6, no. 1, pp. 416–428, May 2023, doi: 10.31539/INTECOMS.V6I1.5916.
- [8] P. Bidang Komputer Sains dan Pendidikan Informatika, D. Akademi Perekam dan Informasi Kesehatan Iris Padang Jl Gajah Mada No, and S. Barat, "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," *Jurnal Edik Informatika Penelitian Bidang Komputer Sains dan Pendidikan Informatika*, vol. 2, no. 2, pp. 213–219, Feb. 2017, doi: 10.22202/EI.2016.V2I2.1465.
- [9] I. A. Nikmatun and I. Waspada, "IMPLEMENTASI DATA MINING UNTUK KLASIFIKASI MASA STUDI MAHASISWA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR," *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, vol. 10, no. 2, pp. 421–432, Nov. 2019, doi: 10.24176/SIMET.V10I2.2882.

- [10] M. Mustafa, "Diabetes prediction dataset," Kaggle. Accessed: Jun. 01, 2024. [Online]. Available: <https://www.kaggle.com/datasets/iammustafatz/diabetes-prediction-dataset/data>
- [11] A. Asroni, H. Fitri, and E. Prasetyo, "Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik)," *Semesta Teknika*, vol. 21, no. 1, pp. 60–64, May 2018, doi: 10.18196/ST.211211.
- [12] F. Alghifari and D. Juardi, "PENERAPAN DATA MINING PADA PENJUALAN MAKANAN DAN MINUMAN MENGGUNAKAN METODE ALGORITMA NAÏVE BAYES," *JURNAL ILMIAH INFORMATIKA*, vol. 9, no. 02, pp. 75–81, Sep. 2021, doi: 10.33884/JIF.V9I02.3755.
- [13] K. Hamidah, A. Voutama, S. Artikel, K. : Faktor, K. Kdd, and R. Linear, "Analisis Faktor Tingkat Kebahagiaan Negara Menggunakan Data World Happiness Report dengan Metode Regresi Linier.," *Explore IT: Jurnal Keilmuan dan Aplikasi Teknik Informatika*, vol. 15, no. 1, pp. 1–7, Jun. 2023, doi: 10.35891/EXPLORIT.V15I1.3874.
- [14] S. Alam, M. Gito Resmi, N. Masripah, and S. Tinggi Teknologi Wastukencana, "Classification of Covid-19 vaccine data screening with Naive Bayes algorithm using Knowledge Discovery in database method," *Journal of Computer Networks, Architecture and High Performance Computing*, vol. 4, no. 2, pp. 177–185, Jul. 2022, doi: 10.47709/CNAHPC.V4I2.1584.
- [15] E. Saraswati, Y. Umaidah, and A. Voutama, "Penerapan Algoritma Artificial Neural Network untuk Klasifikasi Opini Publik Terhadap Covid-19," *Generation Journal*, vol. 5, no. 2, pp. 109–118, Jul. 2021, doi: 10.29407/GJ.V5I2.16125.
- [16] R. C. Utama, F. Fauziah, and R. T. Komalasari, "Aplikasi Chatbot Berbasis Teks Menggunakan Algoritma Naive Bayes Classifier FAQ GrabAds," *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, vol. 6, no. 1, pp. 90–97, Aug. 2021, doi: 10.30998/STRING.V6I1.9919.