

Implementasi *Orange Data Mining* Untuk Prediksi Penderita Diabetes

Adrian Hartanto¹, Thooyibah. T²
Teknik Informatika, Program Pascasarjana, Universitas Pamulang
e-mail: adrian.hartanto692@gmail.com

Abstrak—Diabetes mellitus adalah salah satu penyakit yang memiliki dampak signifikan pada kesehatan masyarakat. Oleh karena itu, upaya untuk mencegah, mendeteksi, dan mengelola diabetes sangat penting. Dalam konteks ini, penelitian ini bertujuan untuk menginvestigasi implementasi *Orange Data Mining* dalam prediksi penderita diabetes. Metode penelitian ini melibatkan pengumpulan data medis dari sejumlah pasien yang telah didiagnosis menderita diabetes atau tidak. Data tersebut mencakup parameter-parameter seperti tekanan darah dan hasil tes glukosa darah. Data yang diperoleh digunakan sebagai dasar untuk analisis prediktif dengan menggunakan perangkat lunak *Orange Data Mining*. Hasil penelitian menunjukkan bahwa *Orange Data Mining* mampu menghasilkan model prediktif yang akurat dalam memprediksi kemungkinan seseorang menderita diabetes. Berdasarkan parameter-parameter yang digunakan, model ini dapat memberikan informasi yang berharga bagi tenaga medis dalam upaya pencegahan dan pengelolaan diabetes. Selain itu, model ini juga dapat digunakan dalam upaya penyaringan untuk mengidentifikasi individu yang berisiko tinggi terkena diabetes, sehingga tindakan pencegahan dapat diambil lebih awal. Di masa depan, penelitian lebih lanjut dan pengembangan model prediktif yang lebih canggih dapat menjadi langkah berikutnya untuk mengoptimalkan peran analisis data dalam penanganan diabetes dan penyakit kronis lainnya.

Kata Kunci— *Machine Learning; Data Mining; Orange Tools; Prediksi; Diabetes.*

I. PENDAHULUAN

Diabetes Mellitus merupakan masalah metabolisme dengan gula darah sangat tinggi. Gejala gula darah tinggi antara lain buang air kecil berlebihan, berat badan turun, penglihatan tidak jelas, dan luka sulit sembuh bahkan semakin parah. Apabila tidak segera diobati, diabetes bisa mengakibatkan masalah kesehatan lain yang serius hingga kematian. Hal ini bisa menyebabkan komplikasi seperti stroke, komplikasi mata, kardiovaskular, gagal ginjal, dan lain-lain [1]. Prediksi merupakan suatu cara yang digunakan untuk mencari kemungkinan di masa depan. Pada teknologi data mining, berdasarkan bukti dan temuan pada data yang digunakan metode peramalan dapat memprediksi kemungkinan pada masa datang. Perhitungan statistik dan matematika sangat berperan dalam melakukan prediksi [2]. Penelitian terdahulu yang berkaitan dengan penggunaan metode *K-NN* dan *Naïve Bayes* digunakan untuk klasifikasi kelulusan mahasiswa [3], prediksi penyakit diabetes [4], prediksi harga USDT/BIDR [5], prediksi harga *bitcoin* [6], prediksi harga jual tanah, seleksi penerima beasiswa. Adapun perbedaan penelitian kali ini adalah tidak hanya membandingkan antara satu metode dengan metode lain dalam hal prediksi, tetapi juga digunakan untuk memprediksi data testing setelah mendapatkan metode mana yang terbaik.

II. METODE PENELITIAN

Tujuan penelitian ini untuk menganalisa perbandingan metode *Naive Bayes* dan *K-NN* untuk memprediksi penderita diabetes menggunakan *Orange Data Mining Tools*.

K-Nearest Neighbor (K-NN) merupakan Algoritma yang digunakan untuk mengklasifikasikan data berdasarkan *data training* yang didapat dari tetangga terdekat, dimana *k* adalah bilangan tetangga terdekat [7]. Ada beberapa tahap kerja *K-NN*, pertama menentukan nilai *K*, kedua menghitung jarak data dengan *K* yang ditentukan dengan *euclidean distance*, ketiga mengambil *K* terdekat untuk dibuat prediksi berdasarkan hasil perhitungan jarak.

Naive bayes adalah klasifikasi dengan suatu metode kemungkinan dan perhitungan yang ditemukan oleh seseorang ilmuwan dari Inggris yaitu Thomas Bayes menghasilkan prediksi peluang yang akan datang berdasarkan suatu pengalaman sebelumnya [8].

Tahap-tahap penelitian ini terdiri dari pengumpulan data, membuat *data training* dan *data testing*, evaluasi model dan hasil perbandingan metode, yang terakhir melakukan prediksi penderita diabetes.

1. Pengumpulan dataset

Variabel pada dataset ini ada 3 yaitu:

- Glukose: menunjukkan angka kandungan glukosa
- Bloodpressure: menunjukkan angka tekanan darah
- Diabetes: menunjukkan seseorang termasuk penderita diabetes atau tidak

2. Pembuatan sampling data training

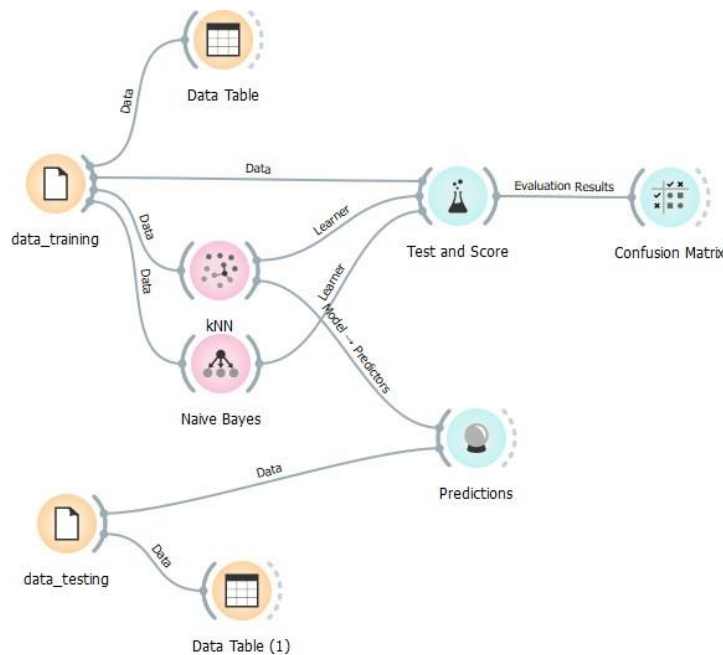
Sampling data training yang digunakan terdapat pada Gambar 1 di bawah ini.

	Glukose	Bloodpressure	Diabetes
1	40	85	No
2	40	92	No
3	45	63	Yes
4	45	80	No
5	40	73	Yes
6	45	82	No
7	40	85	No
8	30	63	Yes
9	65	65	Yes
10	45	82	No
11	35	73	Yes
12	45	90	No
13	50	68	Yes
14	40	93	No
15	35	80	Yes
16	50	70	Yes
17	40	73	Yes
18	40	67	Yes
19	40	75	Yes
20	40	80	Yes
21	40	72	Yes
22	40	88	No

Gambar 1.
 Sampling data training

3. Model prediksi menggunakan Orange *Data Mining Tools*

Model prediksi penderita diabetes bisa dianalisa menggunakan *Orange Data Mining Tools* untuk menentukan metode yang terbaik. Gambar 2 menunjukkan model prediksinya.



Gambar 2.
 Design model prediksi menggunakan Orange Data Mining

4. Evaluasi dan hasil perbandingan dua metode

Setelah dataset dan model prediksi sudah dibuat, maka selanjutnya *data training* diproses ke model prediksi *Orange Data Mining* dengan metode *Naive Bayes* dan *K-NN*.

5. Setelah ditentukan metode terbaik untuk membuat prediksi, maka berikutnya melakukan prediksi dari data testing. Adapun data testing dapat dilihat pada Gambar 3.

	Name	Type	Role	Values
1	Glukose	N numeric	feature	
2	Bloodpressure	N numeric	feature	

Gambar 3.
Data testing

III. HASIL DAN PEMBAHASAN

Data training yang sudah diproses, selanjutnya dilakukan pengujian untuk mendapatkan model prediksi terbaik.

		Predicted		
		No	Yes	Σ
Actual	No	99	9	108
	Yes	7	85	92
Σ		106	94	200

Gambar 4.
Confusion matrix K-NN

Model	AUC	CA	F1	Prec	Recall	MCC
kNN	0.967	0.965	0.965	0.965	0.965	0.930
Naive Bayes	0.980	0.920	0.920	0.920	0.920	0.839

Gambar 5.
Hasil Test and Score

Penelitian kali ini hanya menggunakan nilai *F1 Score* dan *MCC* dimana nilai terbaik *F1* adalah 1.0 dan nilai terburuknya adalah 0. Jika *F1* menunjukkan skor yang baik berarti model klasifikasi memiliki *precision* dan *recall* yang baik. Sedangkan untuk *MCC* memiliki nilai antara -1 hingga 1 dimana -1 menunjukkan klasifikasi biner yang sepenuhnya salah sedangkan 1 menunjukkan klasifikasi biner yang sebenarnya [9]. Gambar 4 menunjukkan perbandingan hasil pengujian dari metode *K-NN* dan *Naive Bayes* menggunakan *widget Test and Score*. Hasil dari *confusion matrix* juga menunjukkan bahwa metode *K-NN* baik untuk melakukan prediksi, ditandai dengan angka *false* yang kecil. Dari beberapa hasil test tersebut maka metode yang terbaik adalah *K-NN* dengan nilai *F1 Score* 0.965 dan nilai *MCC* 0.930.

	Glukose	Bloodpressure
1	40	63
2	40	90
3	50	90
4	25	67
5	55	67

Gambar 6.
Data testing sebelum dilakukan prediksi

kNN	Glukose	Bloodpressure
1 Yes	40	63
2 No	40	90
3 No	50	90
4 Yes	25	67
5 Yes	55	67

Gambar 7.
Hasil prediksi data testing menggunakan K-NN

IV. KESIMPULAN

Hasil implementasi *Orange Data Mining* dalam memprediksi penderita diabetes diperoleh hasil nilai *F1 Score* dan *MCC* dari metode Naive Bayes 0.920 dan 0.839 dan hasil nilai *F1 Score* dan *MCC* dari metode *K-NN* 0.965 dan 0.930. Berdasarkan perbandingan nilai *F1 Score* dan *MCC* tersebut dapat disimpulkan bahwa metode *K-NN* lebih baik untuk digunakan dalam memprediksi penderita diabetes. Adapun hasil prediksinya menunjukkan dari 5 data terdapat 2 orang yang diprediksi tidak mengidap diabetes dan 3 orang yang diprediksi menderita diabetes. Penelitian yang akan datang disarankan bisa membandingkan metode *K-NN* dengan metode lain dalam melakukan prediksi menggunakan *Orange Data Mining* atau *tools* yang lain.

DAFTAR PUSTAKA

- [1] Tigga, N. P., & Garg, S. (2020). Prediction of Type 2 Diabetes using Machine Learning Classification Methods. *Procedia Computer Science*, 167, 706– 716. <https://doi.org/10.1016/j.procs.2020.03.336>
- [2] Suryanto, A. A., Muqtadir, A., & Artikel, S. (2019). *PENERAPAN METODE MEAN ABSOLUTE ERROR (MEA) DALAM ALGORITMA REGRESI LINEAR UNTUK PREDIKSI PRODUKSI PADI* *Info Artikel : ABSTRAK*. 1, 11.
- [3] H. Hozairi, A. Anwari, and S. Alim, "IMPLEMENTASI ORANGE DATA MINING UNTUK KLASIFIKASI KELULUSAN MAHASISWA DENGAN MODEL K-NEAREST NEIGHBOR, DECISION TREE SERTA NAIVE BAYES," *Netw. Eng. Res. Oper. Vol 6, No 2 NERO*, Nov. 2021, [Online]. Available: <https://nero.trunojoyo.ac.id/index.php/nero/article/view/237>.
- [4] Maulidah, N., Supriyadi, R., Utami, D. Y., Hasan, F. N., Fauzi, A., & Christian, A. (2021). Prediksi Penyakit Diabetes Melitus Menggunakan Metode Support Vector Machine dan Naive Bayes. *Indonesian Journal on Software Engineering (IJSE)*, 7(1), 63–68. <http://ejournal.bsi.ac.id/ejurnal/index.php/ijse63>
- [5] Muharrom, M. (2023). Analisis Penggunaan Orange Data Mining untuk Prediksi Harga USDT/BIDR Binance. *Bulletin of Information Technology (BIT)*, 4(2), 178–184. <https://doi.org/10.47065/bit.v3i1>
- [6] Ichsan, N., Fatah, H., Wahyuni, T., & Ermawati, E. (2022). IMPLEMENTASI ORANGE DATA MINING UNTUK PREDIKSI HARGA BITCOIN. *JURNAL RESPONSIF*, 4(2), 118–125. <https://investing.com/crypto/bitcoin/historical->
- [7] Said, H., Matondang, N., Nurramdhani Irmanda, H., & Informasi, S. (2022). *Penerapan Algoritma K-Nearest Neighbor Untuk Memprediksi Kualitas Air Yang Dapat Dikonsumsi* *Application of K-Nearest Neighbor Algorithm to Predict Consumable Water Quality* (Vol. 21, Issue 2). www.kaggle.com
- [8] Rachman, R., Handayani, R. N., & Artikel, I. (2021). Klasifikasi Algoritma Naive Bayes Dalam Memprediksi Tingkat Kelancaran Pembayaran Sewa Teras UMKM. *JURNAL INFORMATIKA*, 8(2). <http://ejournal.bsi.ac.id/ejurnal/index.php/ji>
- [9] Hasdyna, N., & Dinata, R. K. (2020). Analisis Matthew Correlation Coefficient pada K-Nearest Neighbor dalam Klasifikasi Ikan Hias. *INFORMAL: Informatics Journal*, 5(2), 57. <https://doi.org/10.19184/isj.v5i2.18907>