

Penerapan Algoritma Neural Network Untuk Mendeteksi Penyakit Diabetes Berbasis Algoritma Genetika

Amin Rais

Teknik Elektro, Universitas Trunojoyo Madura, Bangakalan
akur5008@gmail.com

Abstract – Diabetes is a disease in which the glucose content in the blood can't patients, first action is needed as a solution of the problem of diabetes by predicting to detect early diabetes. This is necessary because often medical decisions are made based on experience and rational reasoning. Prediction of diabetes can be done by using some data of diabetic patients who have been stored in the database to make a pattern for the determination of diabetes with Artificial Intelligen technique so that the result of inaccurate diagnosis can be avoided. In this study the authors apply Genetic Algorithm (GA) to optimize Neural Network (NN) model by searching the best parameter value on Neural Network (NN) model. The experimental results showed a decrease in RMSE value which means an increase in predicted accuracy value, ie from 0.402 +/- 0,035 to 0.396 +/- 0,032. The optimization of the NN model prediction, the relevant policy makers can determine the prediction of diabetes more accurately.

Keywords — *Diabetes, Artificial Intelligen, Algoritma Genetika, Neural Network, RMSE.*

Abstrak – Diabetes adalah sebuah penyakit dimana kandungan gula dalam darah tidak dapat diolah oleh tubuh. Berdasarkan kasus banyaknya penderita penyakit diabetes, maka perlu tindakan awal sebagai solusi dari permasalahan penyakit diabetes yaitu dengan melakukan prediksi guna mendeteksi dini penyakit diabetes. Hali ini perlu dilakukan karena sering kali keputusan medis yang dibuat berdasarkan pengalaman dan penalaran rasional. Prediksi penyakit diabetes dapat dilakukan dengan cara memanfaatkan beberapa data pasien penderita penyakit diabetes yang telah tersimpan dalam basis data guna dibuat suatu pola untuk penentuan penyakit diabetes dengan teknik Artificial Intelligen sehingga hasil diagnose yang tidak tepat dapat dihindari. Dalam penelitian ini menerapkan Algoritma genetika (GA) untuk mengoptimalkan model Neural Network (NN) dengan cara pencarian nilai parameter terbaik pada model Neural Network (NN). Hasil eksperimen menunjukkan terjadinya penurunan nilai RMSE yang berarti terjadi peningkatan pada nilai akurasi prediksi, yaitu dari 0,402 +/- 0,035 menjadi 0,396 +/- 0,032. Dengan adanya optimasi prediksi model NN, maka para pengambil kebijakan terkait dapat menentukan prediksi penyakit diabetes dengan lebih akurat.

Kata Kunci—*Diabetes, Artificial Intelligen, Algoritma Genetika, Neural Network, RMSE.*

I. PENDAHULUAN

Diabetes atau istilah lainnya Diabetes Melitus (DM) adalah sebuah penyakit dimana kandungan gula dalam darah tidak

dapat diolah oleh tubuh[1]. Kondisi normal tubuh seseorang secara otomatis akan menghasilkan insulin yang berfungsi untuk memastikan kadar gula tetap berada dalam ambang batas yang normal. Ketika tubuh menghasilkan hormon insulin yang tidak cukup atau bahkan sel tidak merespon insulin yang dihasilkan, maka gula darah menjadi tinggi sehingga penyakit diabetes terjadi. Adapun gejala klasik yang terjadi secara umum adalah hasrat untuk buang air kecil yang terlalu sering, rasa haus yang tidak berhenti, dan tingginya ritme rasa lapar[2].

Dalam kurun waktu lebih kurang dari 30 tahun terakhir jumlah penderita penyakit diabetes di Negara Cina mengalami peningkatan yang signifikan, hampir 110 juta jiwa pada tahun 2017. Data ini menunjukkan bahwa Cina memiliki populasi pasien dengan penyakit diabetes terbesar di dunia. Penderita diabetes dengan jenis kelamin pria cenderung lebih tinggi dibandingkan dengan wanita. Hal ini menyebabkan perubahan pola dan gaya hidup warga sekitar yang ingin menjadi lebih sehat [3].

Menurut International Diabetes Federation (IDF), Indonesia menempati peringkat ke enam di dunia dengan penderita DM terbanyak. Berdasarkan data dari Riskesdas Provinsi Jawa Barat , prevalensi diabetes berdasarkan diagnosa nakes dan diagnosis disertai gejala masing-masing sebesar 1,3% dan 2,0%. Populasi pasien penderita penyakit diabetes akan semakin meningkat hingga 642 juta jiwa, yang berarti satu dari sepuluh orang dewasa. Prevalensi untuk Provinsi Jawa Tengah sebesar (1,9%) (Kemenkes RI, 2014). Jumlah kasus DM tipe 2 di Jawa Tengah tahun 2015 sebanyak 99.646 kasus. Hal ini berbeda dengan tiga tahun sebelumnya[4].

Berdasarkan kasus banyaknya penderita penyakit diabetes, maka perlu tindakan awal sebagai solusi dari permasalahan penyakit diabetes yaitu dengan melakukan prediksi guna mendeteksi penyakit diabetes. Hal ini perlu dilakukan karena seringkali keputusan medis yang di buat berdasarkan pengalaman dan penalaran rasional[5]. Oleh karena itu petugas medis dalam hal ini dokter membutuhkan sebuah model untuk memprediksi penyakit diabetes yang handal. Adapun rekomendasi prediksi dari penyakit diabetes ini dapat dibuat berdasarkan catatan sejarah medis pasien penderita diabetes[7]. Prediksi penyakit diabetes dapat dilakukan dengan cara memanfaatkan beberapa data pasien penderita penyakit diabetes yang telah tersimpan dalam basis data guna dibuat suatu pola untuk penentuan penyakit

diabetes dengan teknik Artificial Intelligence sehingga hasil diagnosa yang tidak tepat dapat dihindari [5]. Sebuah kontribusi yang baik dalam pengambilan keputusan dan perencanaan dimasa yang akan datang akan dihasilkan dari prediksi yang akurat[8].

Model prediksi berbagai penyakit telah dilakukan oleh peneliti sebelumnya yakni, penelitian yang dilakukan oleh K. Rajesh, V. Sangeetha (2012) [6], menggunakan data Pima Indian Diabetes Database untuk membandingkan hasil dari beberapa metode diantaranya CRT, CS-RT, C4.5, K-NN, LDA, NAIVE BAYES, PLS-DA, SVM, RND TREE. Diantara 10 metode yang diusulkan algoritma RND TREE merupakan algoritma yang memiliki akurasi paling tinggi yaitu 100%. Namun algoritma tersebut memiliki kelemahan yaitu memiliki rule data yang banyak sehingga penulis memilih C4.5 untuk penelitiannya dimana algoritma C4.5 dinilai lebih efisien. Tingkat akurasi yang dihasilkan mencapai 91%.

A, Andriani (2013) dalam penelitiannya mengklasifikasikan data dari pasien penyakit diabetes [7]. *Confusion matrix* dan kurva ROC digunakan dalam evaluasi hasil klasifikasi tersebut. Dari hasil evaluasi didapatkan hasil *Excellent Clasification. Rule* hasil klasifikasi kemudian diimplementasikan pada sebuah system.

Fatmawati (2016) menerapkan teknik klasifikasi data mining dengan metode Naive Bayes dan Decision Tree [8]. Evaluasi dilakukan menggunakan Confusion Matrix dan kurva ROC. Dari eksperimen yang telah dilakukan Naive Bayes memiliki tingkat akurasi yang lebih baik dari pada Decision Tree dalam memprediksi penyakit diabetes.

II. METODE PENELITIAN

A. Metode

1. Dataset

Dalam penelitian ini dataset yang digunakan adalah dataset publik Pima Indian Diabetes Database, diperoleh dari UCI Machine Learning Repository. Data yang ada pada database berjumlah 768 record yang terdiri dari 9 atribut. Terdiri dari 1 label atau kelas dan 8 atribut lainnya yakni Pregnant, Glucose, DBP, TSFT, INS, BMI, DPF dan Age. Adapun dataset *real* tampak pada Tabel I.

2. Preprocessing Data

Preprocessing data dilakukan sebelum dataset diterapkan kedalam model yang diusulkan yakni dengan melakukan beberapa hal diantaranya :

a. Pemilihan Variabel yang sesuai:

Variabel yang digunakan dataset ini adalah:

- Jumlah kehamilan
- Konsentrasi Glukosa plasma
- Tekanan darah diastolic

- Ketebalan triceps pada lipatan kulit
- Serum insuli 2 jam
- Berat badan
- Silsilah diabetes
- Umur
- Atribut output yaitu keterangan diabetes ‘ya’ atau “tidak” dalam bentuk biner 1 atau 0 yang artinya jika 1 adalah positif diabetes dan 0 negatif.

b. Penentuan Data input dan output

Berdasarkan gejala dan atribut yang tersedia pada dataset untuk mendiagnosa diabetes, maka variable input dan output dataset sebagai berikut :

- Jumlah Kehamilan menggunakan atribut numerik
- Konsentrasi glukosa plasma menggunakan atribut numerik
- Tekanan darah diastolic (mm Hg) menggunakan atribut numerik
- Ketebalan triceps pada lipatan kulit bawah (mm) menggunakan atribut numerik
- Serum insulin 2 jam (mu U/ml) menggunakan atribut numerik
- Berat badan menggunakan (kg) atribut numerik
- Silsilah diabetes menggunakan atribut numerik
- Umur
- Output klasifikasi dengan atribut diabetes (positif/negatif).

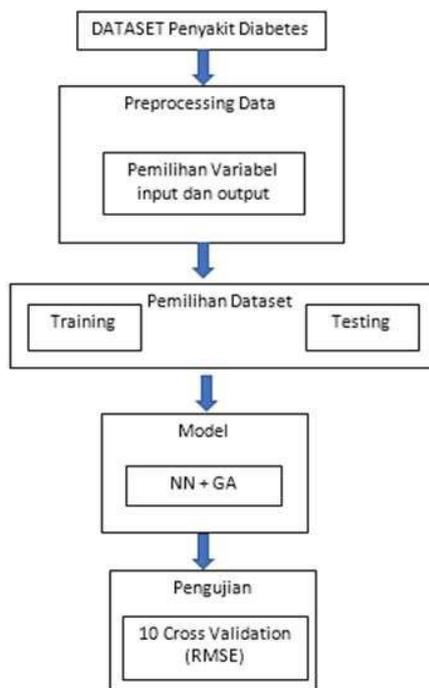
3. Penentuan data training dan testing

Data training diperoleh dari Pima Indian Diabetes Database, diperoleh dari UCI Machine Learning Repository) yang berjumlah 768 record. Dataset kemudian dibagi menjadi dua bagian dengan perbandingan 9:1, yakni 9 bagian data training dan 1 bagian untuk testing.

4. Metode yang diusulkan

Algoritma Genetika (GA) diusulkan untuk dapat meningkatkan tingkat akurasi prediksi awal penyakit diabetes. Pencarian nilai parameter terbaik dilakukan oleh algoritma genetika (GA) yang nantinya akan diimplementasikan pada algoritma Neural Network (NN) sehingga akurasi yang dihasilkan menjadi lebih optimal. Semakin kecil nilai Root Mean Square Error (RMSE) yang dihasilkan maka kualitas prediksi dikategorikan baik. Adapun tools yang digunakan untuk mendapatkan model yang diusulkan adalah RapidMinerStudio8. Sedangkan proses validasi model digunakan K-Fold Validation dengan jumlah k=10. Gbr. 1 menunjukkan alur metode dari model yang diusulkan dalam penelitian.

B. Gambar dan Tabel



Tabel 1
 Data Real

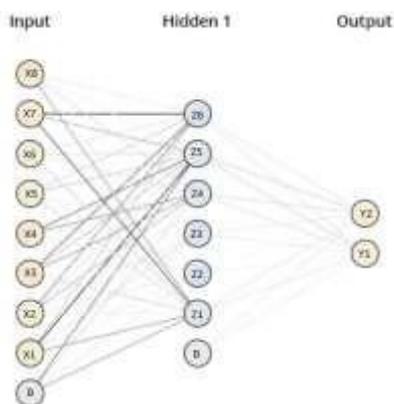
No	Pregnant	Glucose	DBP	TSFT	INS	BMI	DPF	Age	CLASS
1	6	148	72	35	0	33,6	0,627	50	tested_positive
2	1	85	66	29	0	26,6	0,351	31	tested_negative
3	8	183	64	0	0	23,3	0,672	32	tested_positive
4	1	89	66	23	94	28,1	0,167	21	tested_negative
5	0	137	40	35	168	43,1	2.288	33	tested_positive
6	5	116	74	0	0	25,6	0,201	30	tested_negative
7	3	78	50	32	88	31	0,248	26	tested_positive
8	10	115	0	0	0	35,3	0,134	29	tested_negative
9	2	197	70	45	543	30,5	0,158	53	tested_positive
10	8	125	96	0	0	0	0,232	54	tested_positive
11	4	110	92	0	0	37,6	0,191	30	tested_negative
12	10	168	74	0	0	38	0,537	34	tested_positive
13	10	139	80	0	0	27,1	1.441	57	tested_negative
...
...
...
762	2	88	58	26	16	28,4	0,766	22	tested_negative
762	9	170	74	31	0	44	0,403	43	tested_positive
763	9	89	62	0	0	22,5	0,142	33	tested_negative
764	10	101	76	48	180	32,9	0,171	63	tested_negative
765	2	122	70	27	0	36,8	0,34	27	tested_negative
766	5	121	72	23	112	26,2	0,245	30	tested_negative
767	1	126	60	0	0	30,1	0,349	47	tested_positive
768	1	93	70	31	0	30,4	0,315	23	tested_negative

III. HASIL DAN PEMBAHASAN

Eksperimen ini menggunakan komputer dengan spesifikasi processor CPU Intel Core i3 2.10 GHz, RAM 4 GB, serta sistem operasi Windows 7 Ultimate 64-bit. Setelah eksperimen dilakukan, didapatkan hasil yang berkaitan dengan model yang telah diusulkan. Eksperimen dilakukan melalui dua tahapan, yaitu (1) Penerapan Algoritma Neural Network (NN); yaitu dengan melakukan eksperimen parameter Neural Network (NN) terbaik. (2) penerapan Neural Network (NN) dan Algoritma Genetika (GA); yaitu dengan pencarian parameter terbaik Neural Network (NN) terbaik berbasis GA.

A. Hasil Eksperimen Neural Network (NN) Terbaik

Eksperimen yang telah dilakukan dengan menerapkan algoritma Neural Network (NN) terhadap dataset yang telah ditentukan sebagai data training diperoleh arsitektur model seperti gambar dibawah ini:



Gbr. 2 Arsitektur model *Neural Network (NN)* terbaik.

Gambar 2 menunjukkan bahwa data *input* yang digunakan sebanyak delapan *neuron input* sedangkan *output* dihasilkan dua *neuron* serta dihasilkan juga enam *neuron hidden layer*. Pada eksperimen ini ditetapkan nilai parameter yang terbaik adalah *Training Cycles* = 100, *Learning rate* = 0.1, dan *Momentum* = 0.1 sehingga *RMSE* = 0.402 +/- 0.035.

B. Hasil Eksperimen Neural Network (NN) dan GA

Proses pengujian guna mendapatkan nilai parameter terbaik dari Neural Network (NN) dilakukan dengan setting parameter algoritma genetika yakni *Population size* = 5, *Max generations* = 50, *Mutation type* = *Gaussian_mutation*, *Crossover prob* = 0.9 dan *Selection type* = *Roulette wheel*

Hasil Eksperimen NN dan GA

Parameter	Algoritma NN + Algoritma Genetika
Learning rate	0,1985425571526106
Momentum	0,39084178927133306
RMSE	0,396 +/- 0,032

Berdasarkan pengamatan terhadap eksperimen yang telah dilakukan, dengan adanya optimasi pada nilai parameter Neural Network (NN) dengan algoritma genetika (GA) mengalami peningkatan akurasi menjadi 0,396 +/- 0,032. Dengan nilai parameter terbaik untuk *Learning rate* sebesar 0,1985425571526106 dan *Momentum* terbaik sebesar 0,39084178927133306.

C. Evaluasi

Dalam penelitian ini Algoritma Genetika (GA) diterapkan kedalam model Neural Network (NN) dengan tujuan untuk mengoptimalkan tingkat akurasi prediksi deteksi awal penyakit diabetes. Adapun teknik optimasi dilakukan dengan cara pencarian nilai parameter terbaik model Neural Network (NN). Hasil eksperimen menunjukkan terjadinya penurunan nilai *RMSE* yang berarti terjadi peningkatan pada nilai akurasi prediksi, yaitu dari 0,402 +/- 0,035 menjadi 0,396 +/- 0,032

PERBANDINGAN NN DAN NN+GA

Model	RMSE
NN	0,402 +/- 0,035
NN+GA	0,396 +/- 0,032

IV. KESIMPULAN

Eksperimen yang telah dilakukan dengan mengusulkan Algoritma Genetika (GA) untuk diterapkan kedalam model yang digunakan yakni Neural Network (NN) dalam memprediksi deteksi awal penyakit diabetes, dapat disimpulkan bahwa Algoritma Genetika (GA) dapat mengoptimalkan model Neural Network (NN) dengan cara pencarian nilai parameter terbaik pada model Neural Network (NN). Sehingga terjadi peningkatan akurasi prediksi. Dengan adanya optimasi prediksi model tersebut, maka para pengambil kebijakan terkait dapat menentukan prediksi penyakit diabetes dengan lebih tepat. Eksperimen dalam penelitian ini belum cukup untuk dapat menentukan ketepatan akurasi prediksi. Pada penelitian selanjutnya diharapkan untuk dapat menerapkan algoritma optimasi lain guna mengetahui perbedaan tingkat akurasi yang terjadi sehingga nilai *RMSE* menjadi lebih optimal.

V. DAFTAR PUSTAKA

- [1] A. Zamora-Kapoor, A. Fyfe-Johnson, A. Omidpanah, D. Buchwald, and K. Sinclair, "Risk factors for pre-diabetes and diabetes in adolescence and their variability by race and ethnicity," *Prev. Med. (Baltim)*, vol. 115, no. August, pp. 47–52, 2018.
- [2] A. A. Aljumah, M. G. Ahamad, and M. K. Siddiqui, "Application of data mining: Diabetes health care in young and old patients," *J. King Saud Univ. - Comput.*

- Inf. Sci.*, vol. 25, no. 2, pp. 127–136, 2012.
- [3] H. Wu, S. Yang, Z. Huang, J. He, and X. Wang, “Type 2 diabetes mellitus prediction model based on data mining,” *Informatics Med. Unlocked*, vol. 10, no. August 2017, pp. 100–107, 2018.
- [4] R. Premunendar, I. Dewi, and H. Asari, “Penentuan Prediksi Awal Penyakit Jantung Menggunakan Algoritma Back Propagation Neural Network dengan Metode Adaboost,” *Semantik*, vol. 2013, no. November, pp. 298–304, 2013.
- [5] K. Rajesh and V. Sangeetha, “Application of Data Mining Methods and Techniques for Diabetes Diagnosis,” (*sem qualis*) *Int. J. Eng. Res. Innov. Technol.*, vol. 2, no. 3, pp. 224–229, 2012.
- [6] A. Andriani, “Sistem prediksi penyakit diabetes berbasis decision tree,” *J. Bianglala Inform.*, vol. I, no. 1, pp. 1–10, 2013.
- [7] S. Kang, “Personalized prediction of drug efficacy for diabetes treatment via patient-level sequential modeling with neural networks,” *Artif. Intell. Med.*, vol. 85, pp. 1–6, 2018.
- [8] F. Fatmawati, “PERBANDINGAN ALGORITMA KLASIFIKASI DATA MINING MODEL C4.5 DAN NAIVE BAYES UNTUK PREDIKSI PENYAKIT DIABETES,” *J. Techno Nusa Mandiri*, vol. 13, no. 1, pp. 50–59, Mar. 2016.