



@Hak cipta milik IPB University

IPB University

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

PERBANDINGAN KINERJA MODEL *LOGISTIC REGRESSION* DAN *ARTIFICIAL NEURAL NETWORK* DALAM MENDIAGNOSIS PASIEN DIABETES

DESI NOVIYANTI



DEPARTEMEN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2024

ABSTRAK

DESI NOVIYANTI. Perbandingan Kinerja Model *Logistic Regression* dan *Artificial Neural Network* dalam Mendiagnosis Pasien Diabetes. Dibimbing oleh MOCHAMAD TITO JULIANTO dan I WAYAN MANGKU.

Diabetes melitus merupakan kondisi yang terjadi akibat kekurangan produksi insulin, baik sebagian maupun keseluruhan, yang menyebabkan peningkatan kadar glukosa dalam darah. Mendeteksi diabetes sejak dini sangat penting untuk mencegah berkembangnya komplikasi yang serius. Penelitian ini bertujuan untuk membandingkan kinerja model *Logistic Regression* dan *Artificial Neural Network* (ANN) dalam mendiagnosis pasien diabetes. Data penelitian terdiri dari 8 variabel bebas dan 1 variabel target. Hasil penelitian menunjukkan bahwa model ANN merupakan model terbaik berdasarkan hasil metrik evaluasi akurasi dan nilai AUC, baik pada data *training* (akurasi = 81.82% dan AUC = 89.53%) maupun data *testing* (akurasi = 80.73% dan AUC = 87.53%). Sementara itu, model *Logistic Regression* merupakan model terbaik berdasarkan kecepatan waktu komputasinya, baik pada data *training* (0.0010 detik) maupun data *testing* (0.0003 detik).

Kata kunci: artificial neural network, diabetes melitus, *logistic regression*, *machine learning*.

ABSTRACT

DESI NOVIYANTI. Performance Comparison of Logistic Regression and Artificial Neural Network Models in Diagnosing Diabetic Patients. Supervised by MOCHAMAD TITO JULIANTO and I WAYAN MANGKU.

Diabetes mellitus is a metabolic disorder characterized by insufficient insulin production, leading to elevated glucose levels in the bloodstream. Early detection of diabetes is crucial to prevent the development of serious complications. This study aims to compare the performance of Logistic Regression and Artificial Neural Network (ANN) models in diagnosing diabetes patients. The dataset consists of eight independent variables and one target variable. The findings reveal that the ANN model outperforms the Logistic Regression model based on accuracy and AUC (Area Under Curve) metrics. The ANN model achieved an accuracy of 81.82% and an AUC of 89.53% on the training data, and an accuracy of 80.73% and an AUC of 87.53% on the testing data. Conversely, the Logistic Regression model excelled in computational speed, with processing times of 0.0010 seconds for the training data and 0.0003 seconds for the testing data.

Keywords: artificial neural network, diabetes mellitus, *logistic regression*, *machine learning*.



- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Hak Cipta milik IPB, tahun 2024
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.



PERBANDINGAN KINERJA MODEL *LOGISTIC REGRESSION* DAN *ARTIFICIAL NEURAL NETWORK* DALAM MENDIAGNOSIS PASIEN DIABETES

DESI NOVIYANTI

Skripsi
sebagai salah satu syarat untuk memperoleh gelar
Sarjana Matematika pada
Program Studi Matematika

**DEPARTEMEN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2024**



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Penguji pada Ujian Skripsi:
Elis Khatizah S.Si., M.Si., Ph.D.



Judul Skripsi : Perbandingan Kinerja Model *Logistic Regression* dan *Artificial Neural Network* dalam Mendiagnosis Pasien Dsiabetes
Nama : Desi Noviyanti
NIM : G5401201022

Disetujui oleh

Pembimbing 1:
Mochamad Tito Julianto, S.Si., M.Kom.

Pembimbing 2:
Prof. Dr. Ir. I Wayan Mangku, M.Sc.

Diketahui oleh

Ketua Departemen Matematika:
Prof. Dr. Ir. Endar Hasafah Nugrahani, M.S.
NIP 19631228 198903 2 001

Tanggal Ujian: 26 November 2024

Tanggal Lulus:



PRAKATA

Puji dan syukur penulis panjatkan kepada Allah subhanaahu wa ta'ala atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Judul yang dipilih dalam penelitian ini dilaksanakan sejak bulan Oktober 2023 sampai bulan Oktober 2024 adalah “Perbandingan Kinerja Model *Logistic Regression* dan *Artificial Neural Network* dalam Mendiagnosis Pasien Diabetes”.

Pembuatan karya ilmiah ini tidak akan terjadi tanpa bantuan dari seluruh pihak terkait. Terima kasih penulis ucapkan kepada:

1. Bapak Sutrisno dan Ibu Dedeh selaku orang tua penulis, Dudung dan Herlansyah selaku kakak kandung penulis, serta seluruh keluarga besar penulis yang selalu memberikan doa, kasih sayang, dan dukungan penuh, baik moral maupun material, selama penulis menyelesaikan penelitian ini.
2. Bapak Mochamad Tito Julianto, S.Si., M.Kom selaku dosen pembimbing 1, atas segala waktu, bimbingan, dan arahan yang telah diberikan kepada penulis dalam menyelesaikan penelitian ini.
3. Bapak Prof. Dr. Ir. I Wayan Mangku, M.Sc. selaku dosen pembimbing 2, atas segala waktu, bimbingan, dan arahan yang telah diberikan kepada penulis dalam menyelesaikan penelitian ini.
4. Elis Khatizah S.Si., M.Si., Ph.D. selaku dosen penguji yang telah memberikan saran dan masukan kepada penulis dalam penulisan karya ilmiah ini.
5. Seluruh Dosen dan Tenaga Kependidikan Departemen Matematika IPB University atas segala ilmu dan bantuannya selama masa perkuliahan dan penelitian.
6. Desi Noviyanti, selaku penulis, atas ketekunan, dedikasi, dan semangat pantang menyerah yang ditunjukkan sepanjang perjalanan penelitian ini.
7. Siti Nur Lathifah, Nilam Nindya Garini, Ghevira Chairunnisa, Hani Aqila Salsabila, Sherly Yulianty, Rifdah Nur Alifah, Salsabila Fitri Imni, Dita Dervinta, Aqilah Fitri Chania, Renda Sulistyorini, Syammira Dhifa Maulia, Ana Nur Fitria, Alifia Nabila Hakim, Tazkia Nuur Hafizha, Annisa Permata Sari, dan Rizka Dwi Andriani atas segala kebersamaan, ketenangan, dan tempat berbagi cerita dalam menyelesaikan penelitian ini.
8. Teman-teman Departemen Matematika angkatan 57 yang telah menemani perjalanan penulis selama masa studi di IPB University.
9. Pihak-pihak lainnya yang telah membantu penulis yang tidak dapat disebutkan satu per satu.

Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan.

Bogor, Desember 2024

Desi Noviyanti

DAFTAR ISI

DAFTAR TABEL	x
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN	xi
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan	2
II TINJAUAN PUSTAKA	3
2.1 <i>Machine Learning</i>	3
2.2 <i>Missing Values</i>	3
2.3 Standardisasi	4
2.4 <i>Logistic Regression</i>	4
2.5 <i>Artificial Neural Network (ANN)</i>	7
2.6 Evaluasi Model	14
III METODE	16
3.1 Data	16
3.2 Tahapan Penelitian	16
IV HASIL DAN PEMBAHASAN	19
4.1 Pra-pemrosesan Data	19
4.2 Model <i>Machine Learning</i>	28
4.3 Evaluasi Model	35
V SIMPULAN DAN SARAN	37
5.1 Simpulan	37
5.2 Saran	37
DAFTAR PUSTAKA	38
LAMPIRAN	41
RIWAYAT HIDUP	44

DAFTAR TABEL

1	<i>Confusion matrix</i>	14
2	Klasifikasi nilai AUC	15
3	Variabel dalam data Pima Indians Diabetes	16
4	Banyaknya nilai nol pada data	23
5	Banyaknya nilai nol pada data setelah dilakukan penanganan <i>missing value</i>	25
6	Pembagian data Pima Indians Diabetes	28
7	Hasil estimasi koefisien dan <i>odds ratio</i>	28
8	<i>Confusion matriks</i> hasil klasifikasi data <i>training</i>	29
9	<i>Confusion matriks</i> hasil klasifikasi data <i>testing</i>	30
10	Evaluasi kinerja model <i>Logistic Regression</i>	31
11	Kontruksi model <i>Artificial Neural Network</i>	32
12	<i>Confusion matriks</i> hasil klasifikasi data <i>training</i>	33
13	<i>Confusion matriks</i> hasil klasifikasi data <i>testing</i>	33
14	Evaluasi kinerja model <i>Artificial Neural Network</i> (ANN)	34
15	Perbandingan nilai akurasi model <i>Logistic Regression</i> dan <i>Artificial</i>	35

DAFTAR GAMBAR

1	Ilustrasi model <i>Logistic Regression</i>	5
2	Ilustrasi model <i>Feed-Forward Neural Network</i> (FNN)	7
3	Ilustrasi model <i>Backpropagation Neural Network</i> (BNN)	8
4	Arsitektur <i>Artificial Neural Network</i> (ANN):	8
5	Ilustrasi alur kerja <i>Backpropagation Neural Network</i> (BNN)	9
6	Kurva fungsi aktivasi: (a) Sigmoid, (b) ReLU	12
7	Ilustrasi AUC untuk dua model klasifikasi yang berbeda	15
8	Diagram alur penelitian	17
9	Sebaran variabel bebas: (a) <i>Boxplot</i> untuk menunjukkan distribusi data,	19
10	Distribusi <i>Outcome</i>	23
11	Korelasi antar variabel sebelum pra-pemrosesan	24
12	Korelasi antar variabel setelah pra-pemrosesan	26
13	Variabel bebas sebelum dan sesudah dilakukan standardisasi	27
14	Kurva AUC model <i>Logistic Regression</i>	31
15	Grafik hasil evaluasi model <i>Artificial Neural Network</i> (ANN):	34



DAFTAR LAMPIRAN

1	Cuplikan Data Pima Indians Diabetes	42
2	Cuplikan Data Pima Indians Diabetes Yang Sudah Dilakukan	42
3	Sebaran Nilai Variabel Pima Indians Diabetes Sebelum Dilakukan Pra-pemrosesan	42
4	Sebaran Nilai Variabel Pima Indians Diabetes Setelah Dilakukan Pra-pemrosesan	43
5	Code Standardisasi	43
6	Code Imputasi <i>Linear Regression</i>	43



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.