Jurnal Sistim Informasi dan Teknologi
https://jsisfotek.org/index.php
2022 Vol. 4 No. 2 Hal: 82-86 e-ISSN: 2686-3154

Metode Bayesian dan Multilayer Percepton dalam Mengklasifikasi Diabetes Mellitus

Rasna[1]✉, Moh. Rahmat Irjii Matdoan[2]

[1]Universitas Yapis Papua
[2]Universitas Sains dan Teknologi Jayapura

rasna@uniyap.ac.id

Abstract

Diabetes mellitus is a chronic metabolic disorder that causes glucose regulation in the blood. Blood sugar anomalies can be defined as unwanted readings either due to normal causes or reasons unknown to the patient. Machine learning applications have been widely introduced in diabetes research and blood sugar anomaly detection. However, modeling options and strategies for classification in diabetes mellitus are needed. This study aims to classify the data as diabetic or non-diabetic and improve classification accuracy. Classification accuracy is improved by using many of the data sets as training data and test data. Classification accuracy is improved by using multiple of the data set as data. In the test, the C4.5 and RF hybrid methods, as well as the MLP and Net Bayes hybrid classification methods were developed for the classification of diabetes. In the case of C4.5 + RF it provides an accuracy of 79.31% which is higher than the individual models. Similarly, MLP + Net Bayes, provides an 81.89% higher accuracy than the individual models. In the second case the 85-15% ensemble model training and test partitions have an important role for the classification of diabetes data. The proposed MLP + Net Bayes provides 81.89% accuracy as a robust model for data classification. So that the proposed model achieves the highest accuracy of 81.89% with 6 features and reaches the highest sensitivity of 64.10% and the highest specificity of 90.90%.

Keywords: Diabetes, Bayesian, Multilayer, Percepton, Classification.

Abstrak

Diabetes Mellitus merupakan gangguan metabolisme kronis yang menyebabkan regulasi glukosa pada darah. Anomaly gula darah dapat didefinisikan sebagai pembacaan yang tidak diinginkan baik karena penyebab normal atau alasan yang tidak diketahui oleh pasien. Aplikasi pembelajaran mesin telah diperkenalkan secara luas dalam penelitian diabetes dan deteksi anomaly gula darah. Namun, diperlukan opsi pemodelan dan strategi untuk klasifikasi pada diabetes mellitus. Penelitian ini bertujuan mengklasifikasikan data sebagai diabetes atau non diabetes dan meningkatkan akurasi klasifikasi. Akurasi klasifikasi ditingkatkan dengan menggunakan banyak dari kumpulan data sebagai data latih dan data uji. Akurasi klasifikasi ditingkatkan dengan menggunakan banyak dari kumpulan data sebagai data. Dalam pengujian, metode hybrid C4.5 dan RF, serta klasifikasi dengan metode hybrid MLP dan Net Bayes dikembangkan untuk klasifikasi diabetes. Dalam kasus C4.5 + RF memberikan akurasi 79,31% yang lebih tinggi dari model individu. Demikian pula MLP + Net Bayes, memberikan akurasi 81,89% yang lebih tinggi dari model individu. Dalam kasus kedua model ensemble 85-15% partisi latih dan uji memiliki peranan penting untuk klasifikasi data diabetes. MLP + Net Bayes yang diusulkan memberikan akurasi 81,89% sebagai model yang kuat untuk klasifikasi data. Sehingga model yang diusulkan mencapai akurasi tertinggi sebesar 81,89% dengan 6 fitur dan mencapai sensitivitas tertinggi 64,10% dan spesifisitas tertinggi 90,90%.

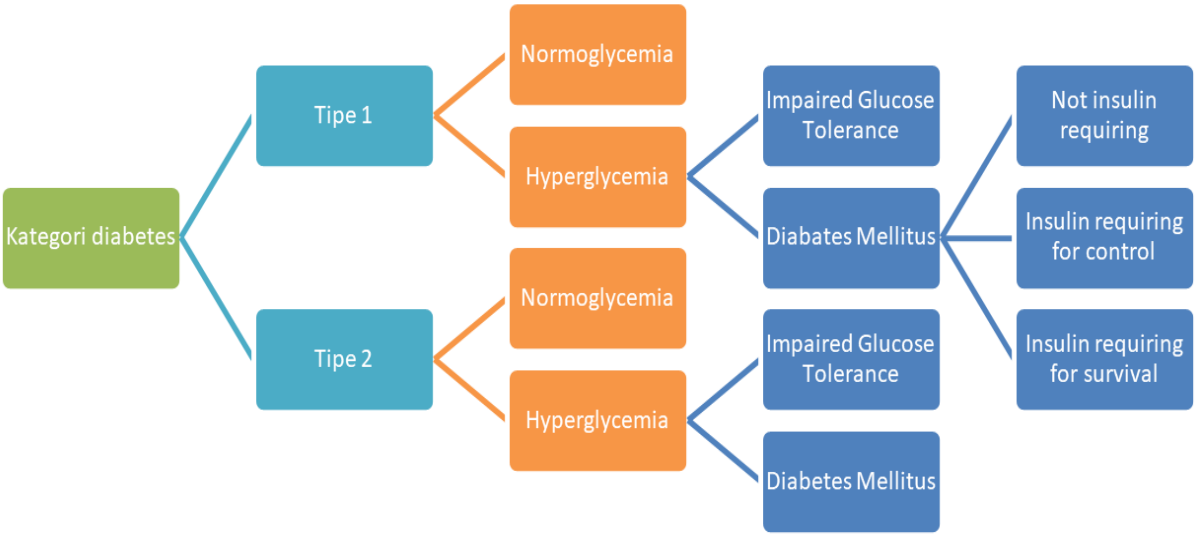
Kata kunci: Diabetes, Bayesian, Multilayer, Percepton, Klasifikasi

1. Pendahuluan

Diabetes Mellitus atau gula darah merupakan gangguan metabolism kronis yang mengakibatkan regulasi gula darah menjadi abnormal [1]. Diperkirakan jumlah penderita diabetes yang berusia antara 20 – 79 tahun adalah 415 juta. Pada tahun 2015, penderita penyakit ini diperkirakan mencapai 642 juta, dan pada tahun 2040 diperkirakan penderita diabeter 521-829 juta. Jumlah kematian yang dikaitkan dengan penyakit diabetes berusia antara 20 – 79 tahun [2], [3].

Penyakit Diabetes Mellitus memiliki berapa tipe, yaitu tipe 1 dan 2 [4]. Kedua tipe ini sangat sulit dibedakan dan hanya dibedakan atas penyebab penyakit tersebut [5]. Tipe 1 disebabkan oleh keturunan dan tipe 2 disebabkan oleh gaya hidup yang kurang sehat [6]. Kedua tipe tersebut disajikan pada Gambar 1.

Diterima: 13-06-2022 | Revisi: 29-06-2022 | Diterbitkan: 30-06-2022 | doi: 10.37034/jsisfotek.v4i2.132



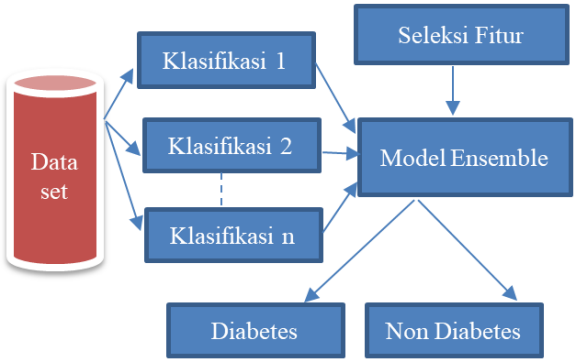
Gambar 1. Tipe Diabetes

Dinamika penyakit gula darah dipengaruhi oleh berbagai factor umum, individu dan tak terduga [7]. Factor umum diantaranya adalah jumlah asupan makanan, asupan insulin, tingkat gula darah sebelumnya, kehamilan, mengkonsumsi obat dan vitamin, merokok dan asupan alcohol. Factor individu seperti beban latihan fisik dan menstruasi. Sedangkan factor tak terduga antara lain stress, penyakit bawaan dan infeksi [8]. Variasi normal penyebab gula darah disebabkan oleh factor-faktor umum dan individu tersebut, sedangkan penyebab gula darah karena factor khusus seperti hipoglikemia dan hiperglikemia tidak dapat diprediksi dan faktor kurangnya pemahaman pasien yang sering menyuntikkan insulin untuk menurunkan kadar gula darah [9].

Dalam pelayanan kesehatan terhadap gula darah terdapat data yang sangat besar dan tidak memiliki nilai pengawasan untuk dijadikan sebagai informasi dan pengetahuan dikarenakan terbatasnya biaya. Diagnosis sangan penting dan menantang untuk diteliti lebih lanjut [10]. Sistem dengan menggunakan data mining dapat membantu pihak yang terlibat dalam dunia kesehatan [11].

Tujuan dalam penelitian ini adalah untuk mendeteksi Diabetes Mellitus menggunakan klasifikasi Bayesian dan Multilayer Perceptron. Kemudian mengklasifikasikan data tersebut menjadi diabetes dan non diabetes. Kategori diabetes dikategorikan menjadi dua kategori. Dalam penelitian ini disebutkan tipe 1 diabetes dan tipe 2 diabetes. Data diklasifikasikan sebagai diabetes atau non diabetes dengan meningkatkan akurasi. Klasifikasi dibutuhkan banyak jumlah sampel yang dipilih untuk menghasilkan akurasi klasifikasi yang lebih tinggi. Untuk mencapai akurasi yang tinggi, maka digunakan banyak kumpulan data sebagai data latih dan data uji.

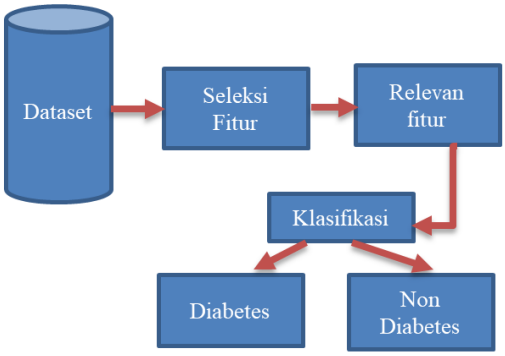
Tahapan penelitian terbagi menjadi tiga yaitu tahap pertama akurasi klasifikasi diperoleh dengan model individu. Tahap kedua, model Ensemble yang digunakan untuk mencapai akurasi tinggi. Tahap ketiga yaitu Teknik feature selection diterapkan pada best ensemble model untuk mencapai akurasi tinggi yaitu efisien secara komputasi. Model sistem yang diusulkan terlihat seperti pada Gambar 2.



Gambar 2. Model Sistem Usulan

2. Metodologi Penelitian

Dalam penelitian kami menggunakan teknik data mining yaitu multi layer perceptron dan klasifikasi Bayesian kemudian menggabungkannya untuk klasifikasi data diabetes. MLP merupakan pengembangan dari perceptron sederhana di lapisan tersembunyi. Dalam proses ini.lebih dari satu lapisan tersembunyi dapat digunakan. Jaringan topologi dibatasi Multilayer Perceptron adalah jaringan saraf yang melatih menggunakan pembelajaran propagasi [12], [13]. Bayesian Net adalah pengklasifikasi statistic yang dapat memprediksi probabilitas keanggotaan, misalnya X adalah sampel data dari label kelas yang tidak diketahui [14]. H menjadi beberapa hipotesis, sehingga sampel data X menjadi milik kelas C tertentu. Untuk masalah klasifikasi, kami mengansumsikan P(H|X) adalah probabilitas, H adalah hipotesis yang berlaku pada sampel data pengamatan yang diberikan oleh X. P(H|X) adalah probabilitas posterior atau probabilitas posteriori dari H dikondisikan pada X. Kemudian dua atau lebih model Bersama-sama membentuk model baru yang disebut dengan model ensemble. Model ensemble adalah kombinasi dari dua atau lebih model untuk menghindari kelemahan dari model individu dan untuk mencapai akurasi yang tinggi. Kemudian dua Teknik dikombinasikan dalam model dan digabungkan ke serangkaian k model yang dipelajari (klasifikasi) M1, M2,…Mk, dengan tujuan membuat model komposit yang ditingkatkan oleh M [15]. Teknik pemilihan fitur untuk memperoleh informasi terlihat seperti pada Gambar 3.



Gambar 3. Fitur Pemilihan Model

2.1. Dataset

Bagian ini berisi kumpulan data dan kinerja ukuran model yang diambil dari dataset Diabetes Phima yang dikumpulkan dari UCI repository kemudian diklasifikasikan dalam dua metode diabetes dan non diabetes. Dataset terdiri dari 8 atribut dan 1 kelas. Dataset terdiri 768 instance yang disajikan pada Tabel 1.

Tabel 1. Dataset Diabetes

Id_Fitur	Nama_Fitur
F1	Pregnant
F2	Plasma Glucose
F3	Diastolic Blood Pressure
F4	Triceps Skin Fold Thickness
F5	Serum Insulin
F6	Vody Mass Index
F7	Diabetes Pedigeree
F8	Age
Kelas	Diabetes atau non diabetes

2.2. Kinerja model

Kinerja model dievaluasi dengan ukuran yaitu akurasi klasifikasi, sensitivitas dan spesifisitas. Tahapan evaluasi menggunakan true positive (TP), true negative (TN), positif palsu (FP), dan negative palsu [16].

Tabel 2. Matrix Konfunsion

Sebenarnya banding Prediksi	Positif	Negatif
Positif	TP	FN
Negatif	FP	TN

Ukuran kinerja seperti sensitivitas, spesifisitas dan akurasi dihitung menggunakan matriks [17] yang disajikan pada Tabel 3.

Tabel 3. Nilai Performa

Akurasi	(TP + TN) / (TP + FP + TN + FN)
Sensitifitas	TP / (TP + FN)
Spesifikasi	TN / (TN + FP)

Peringkat fitur dari yang kurang penting hingga yang sangat penting seperti ditunjukkan F3, F7, F1, F4, F5, F8, F6, F2. Dalam pengujian ini kami menghilangkan fitur dengan memberikan model terbaik yaitu MLP + Bayesian Net [19]. Akurasi yang dicapai sama dengan 6 subset fitur sama seperti di subset fitur lengkap. Ada 8 fitur dalam kumpulan data setelah menghapus fitur F3 dan F7 dari dataset, diperoleh akurasi model 81,69% yang menunjukkan bahwa usulan model memberikan kinerja yang lebih baik dengan jumlah fitur yang lebih sedikit. Pada Tabel 4 menunjukkan berbagai ukuran kinerja model terbaik dan Tabel 5 menunjukkan bahwa akurasi model dengan perbedaan subset fitur.

Tabel 4. Ukuran Kinerja Model Terbaik dengan 6 Fitur

Sebenarnya banding Prediksi	Diabetes	Non Diabetes
Diabetes	17	11
Non Diabetes	10	8

Tabel 5. Seleksi fitur dengan model terbaik

Fitur	Fitur yang dihapus	Nama_Fitur
8	Fitur penuh	81,89
7	F3	81,03
6	F3, F7	81,89
5	F1, F3, F7	76,72
4	F4, F1, F3, F7	80,17
3	F5, F4, F1, F3, F7	81,03
2	F8, F5, F4, F1, F3, F7	77,58
1	F6, F8, F5, F4, F1, F3, F7	70,68

Berdasarkan tabel diatas, model yang diusulkan dapat membantu untuk klasifikasi diabetes.dengan fitur tertinggi yaitu 81,89.

3. Hasil dan Pembahasan

Pengujian dilakukan dengan menggunakan kode Java dan library yang tersedia di WEKA [18]. Dalam pengujian, digunakan berbagai individu dan model klasifikasi hibrida untuk klasifikasi dataset diabetes. Analisis model dilakukan dalam dua tahap yaitu model pertama dilatih dan diuji. Variasi teknik data mining seperti C4.5, random forest (RF), Net Bayes dan Multi layer perceptron (MLP) dilatih menggunakan data latih secara acak kemudian pengujian model dilakukan menggunakan data yang diuji secara acak. Akurasi model dipengaruhi oleh dataset. Akurasi bervariasi, seperti akurasi model C4.5, 77,08% dari data latih dan data uji 75-25%. 76,29% dalam hal 85-15% partisi data latih dan pengujian. 75,32% dalam hal 90-10% partisi latih dan pengujian.

Demikian pula akurasi bervariasi untuk model lain dipartisi yang berbeda, model hybrid akurasinya lebih tinggi dibandingkan model lainnya. Dalam pengujian, metode hybrid C4.5 dan RF, serta klasifikasi dengan metode hybrid MLP dan Net Bayes dikembangkan untuk klasifikasi diabetes. Dalam kasus C4.5 + RF memberikan akurasi 79,31% yang lebih tinggi dari model individu. Demikian pula MLP + Net Bayes, memberikan akurasi 81,89% yang lebih tinggi dari model individu. Dalam kasus kedua model ensemble 85-15% partisi latih dan uji memiliki peranan penting untuk klasifikasi data diabetes. MLP + Net Bayes yang diusulkan memberikan akurasi 81,89% sebagai model yang kuat untuk klasifikasi data, seperti terlihat pada Tabel 5.

Tabel 5. Hasil Klasifikasi Model

Akurasi	81.89%
Spesifikasi	64,10%
Spesifikasi	90,90&

4. Kesimpulan

Dalam penelitian ini, kami telah mengambil berbagai metode klasifikasi dan ansambel untuk memberikan model hybrid dalam pencarian menemukan hasil yang lebih baik dalam hal akurasi, spesifisitas dan sensitifitas. Model yang diusulkan mencapai akurasi tertinggi sebesar 81,89% dengan 6 fitur dan mencapai sensitivitas tertinggi 64,10% dan spesifisitas tertinggi 90,90%.

Daftar Rujukan

[1] Akyol, K., & Şen, B. (2018). Diabetes Mellitus Data Classification by Cascading of Feature Selection Methods and Ensemble Learning Algorithms. International Journal of Modern Education and Computer Science, 10(6), 10–16. doi:10.5815/ijmecs.2018.06.02

[2] Sarki, R., Ahmed, K., Wang, H., Zhang, Y., & Wang, K. (2018). Convolutional Neural Network for Multi-class Classification of Diabetic Eye Disease. ICST Transactions on Scalable Information Systems, 172436. doi:10.4108/eai.16-12-2021.172436

[3] Puchulu, F. (2018). Definition, Classification and Diagnosis of Diabetes Mellitus. Cutaneous Manifestations of Diabetes, 1–1. doi:10.5005/jp/books/13050_2

[4] Alam, U., Asghar, O., Azmi, S., & Malik, R. A. (2014). General aspects of diabetes mellitus. Handbook of clinical neurology, 126, 211-222. DOI: https://doi.org/10.1016/B978-0-444-53480-4.00015-1

[5] Schmidt, M. I., Matos, M. C., Reichelt, A. J., Forti, A. C., De Lima, L., Duncan, B. B., & Group, F. T. B. G. D. S. (2000). Prevalence of gestational diabetes mellitus–do the new WHO criteria make a difference?. Diabetic Medicine, 17(5), 376-380. DOI: https://doi.org/10.1046/j.1464-5491.2000.00257.x

[6] Kanaya, A. M., Grady, D., & Barrett-Connor, E. (2002). Explaining the sex difference in coronary heart disease mortality among patients with type 2 diabetes mellitus: a meta-analysis. Archives of internal medicine, 162(15), 1737-1745. DOI: 10.1001/archinte.162.15.1737

[7] Tri Hastuti, R. (2008). Faktor-faktor Risiko Ulkus Diabetika Pada Penderita Diabetes Mellitus (Studi Kasus di RSUD Dr. Moewardi Surakarta) (Doctoral dissertation, Program Pasca Sarjana Universitas Diponegoro).

[8] Song, S., Zhang, Y., Qiao, X., Duo, Y., Xu, J., Peng, Z., ... & Wang, A. (2022). HOMA-IR as a risk factor of gestational diabetes mellitus and a novel simple surrogate index in early pregnancy. International Journal of Gynecology & Obstetrics, 157(3), 694-701. DOI: https://doi.org/10.1002/ijgo.13905

[9] Entezari, M., Hashemi, D., Taheriazam, A., Zabolian, A., Mohammadi, S., Fakhri, F., ... & Samarghandian, S. (2022). AMPK signaling in diabetes mellitus, insulin resistance and diabetic complications: A pre-clinical and clinical investigation. Biomedicine & Pharmacotherapy, 146, 112563. https://doi.org/10.1016/j.biopha.2021.112563

[10] Anil, K. S., & Jain, R. (2022, April). Data Mining Techniques in Diabetes Prediction and Diagnosis: A Review. In 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI) (pp. 1696-1701). IEEE. DOI: 10.1109/ICOEI53556.2022.9776754

[11] Paisanwarakiat, R., Na-udom, A., & Rungrattanaubol, J. (2022). Combining Logistic Regression Analysis with Data Mining Techniques to Predict Diabetes. In International Conference on Computing and Information Technology (pp. 88-98). Springer, Cham. DOI: https://doi.org/10.1007/978-3-030-99948-3_9

[12] Lagman, A. C., Alfonso, L. P., Goh, M. L. I., Lalata, J. P., Magcuyao, J. P. H., & Vicente, H. N. (2020). Classification Algorithm Accuracy Improvement for Student Graduation Prediction Using Ensemble Model. International Journal of Information and Education Technology, 10(10), 723–727. doi:10.18178/ijiet.2020.10.10.1449

[13] Ramchoun, H., Amine, M., Idrissi, J., Ghanou, Y., & Ettaouil, M. (2016). Multilayer Perceptron: Architecture Optimization and Training. International Journal of Interactive Multimedia and Artificial Intelligence, 4(1), 26. doi:10.9781/ijimai.2016.415

[14] Jackman, S. (2004). Bayesian Analysis for Political Research.

Annual Review of Political Science, 7(1), 483–505. doi:10.1146/annurev.polisci.7.012003.104706

[15] Trust-Region Methods and Conic Model Methods. (n.d.). Springer Optimization and Its Applications, 303–351. doi:10.1007/0-387-24976-1_6

[16] Lagman, A. C., Alfonso, L. P., Goh, M. L. I., Lalata, J. P. & Magcuyao, J. P. H. (2020). Classification Algorithm Accuracy Improvement for Student Graduation Prediction Using Ensemble Model, International Journal of Information and Education Technology, 10(10), 723–727. doi: 10.18178/ijiet.2020.10.10.1449

[17] Tharwat, A. (2020). Classification assessment methods. Applied Computing and Informatics. https://doi.org/10.1016/j.aci.2018.08.003

[18] Russell, I., & Markov, Z. (2017, March). An introduction to the Weka data mining system. In Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education (pp. 742-742). DOI: https://doi.org/10.1145/3017680.3017821