JURNAL PUBLIKASI ILMU KOMPUTER DAN MULTIMEDIA
Halaman Jurnal: http://ejurnal.stie-trianandra.ac.id/index.php/jupkom
Halaman UTAMA Jurnal : http://ejurnal.stie-trianandra.ac.id/index.php

PENERAPAN METODE STACKING DALAM MENGKLASIFIKASIKAN PENDERITA PENYAKIT DIABETES

Binti Mamluatul Karomah
Fakultas Teknik dan Ilmu Komputer / Teknik Informatika, mamluatul93@unusia.ac.id , UNUSIA

ABSTRACT

The International Diabetes Federation (IDF) Atlas 2017 noted that diabetes in Indonesia is increasing. Diabetes mellitus is a disease that has a high complexity, medical care is needed to reduce the impact of complications. which is expected to reduce the risk of complications in diabetic patients in the future. In this research, the Stacking method will be used using C4.5 and SVM as the base model with logistic regression as a meta model in the classification of diabetics. The results indicate the Stacking algorithm improve the performance of the base/single classifier in accuracy, recall and precision. Better stacking method can be use in future research to improve classification performance in diabetics by modifying the base model and meta model.

Keywords: *data mining, stacking, diabetes, ensemble, classification*

Abstrak

International Diabetes Federation (IDF) Atlas 2017 mencatat bahwa penyakit diabetes di Indonesia mengalami peningkatan.Penyakit diabetes melitus adalah penyakit yang memiliki kompleksitas tinggi, perawatan medis yang berkelanjutan sangat dibutuhkan untuk menurunkan dampak dari komplikasi.Dengan adanya dampak yang diakibatkan, maka dapat dilakukan deteksi dini yang diharapkan dapat menurunkan resiko komplikasi pada pasien diabetes diwaktu mendatang. Dalam penelitian ini akan digunakan metode Stacking Stacking menggunakan C4.5 dan SVM sebagai base model dengan logistic regression sebagai meta model dalam klasifikasi penderita penyakit diabetes. Hasil dari penelitian ini menunjukan bahwa algoritma Stacking mampu meningkatkan performa base/single classifier dari sisi accuracy, recall dan precision. Penggunaan metode stacking yang lebih baik dapat dilakukan pada penelitian selanjutnya untuk meningkatkan performa klasifikasi pada penderita diabetes dengan memodifikasi pada base model dan meta model nya.
Kata Kunci: Data Mining, *Stacking*, Penyakit diabetes, *Ensemble*, Klasifikasi

1. PENDAHULUAN

Diabetes melitus merupakan penyakit metabolis kronis yang mana pesien penyakit diabetes tidak menghasilkan jumlah insulin yang cukup atau bisa dikatakan tubuh pasien tidak sanggup memanfaatkan insulin dengan baik sehingga menyebabkan gula darah di dalam tubuh mengalami jumlah yang berlebihan, kondisi ini sering kali dirasakan setelah komplikasi terjadi pada organ tubuh [1]. Pasien didiagnosa menderita penyakit diabetes pada saat kadar glukosa darahnya melebihi nilai normal [2]. Menurut WHO pada tahun 2014 memperkirakan sebanyak 422 juta orang secara global dengan kategori dewasa yang memiliki usia diatas 18 tahun menderita diabetes, dimana diperkirakan wilayah Asia Tenggara dan Pasifik Barat memiliki jumlah terbesar yaitu sekitar 50% kasus diabetes dunia. International Diabetes Federation (IDF) Atlas 2017 mencatat bahwa penyakit diabetes di Indonesia mengalami peningkatan. Hal tersebut didukung oleh Kementrian Kesehatan pada tahun 2018 bahwa Indonesia berada pada peringkat keenam di dunia dengan jumlah penderita diabetes kategori usia 20-79 tahun sekitar 10,3 juta orang, dimana negara-negara seperti Tiongkok, India, Amerika Serikat, Brazil, dan Meksiko memiliki penderita lebih banyak sehingga berada pada peringkat diatasnya [3].

Received Juni 30, 2022; Revised Juli 2, 2022; Accepted Agustus 22, 2022

Penyakit diabetes melitus adalah penyakit yang memiliki kompleksitas tinggi, perawatan medis yang berkelanjutan sangat dibutuhkan untuk menurunkan dampak komplikasi salah satunya dengan pengecekan glikemik [4]. Dampak dari penyakit Diabetes Mellitus yaitu adanya kerusakan vascular mikro seperti retinopati yaitu gangguan pada mata dan neuropati yaitu kerusakan jaringan saraf [5]. Dengan adanya dampak yang diakibatkan oleh penyakit diabetes, maka dapat dilakukan deteksi dini. Deteksi sejak dini diharapkan dapat menurunkan resiko komplikasi pada pasien diabetes diwaktu mendatang.

Penelitian dalam dunia medis untuk memprediksi penyakit sudah banyak dilakukan. Salah satu metode yang paling banyak digunakan yaitu machine learning. Machine learning dapat melakukan pembelajaran dari data sehingga memungkinkan komputer untuk melakukan klasifikasi atau prediksi berdasarkan data tersebut [6]. Penggunaan machine learning pada dunia medis dapat mengurangi biaya dan mempercepat waktu diagnosa beberapa kali lipat. Pada kasus deteksi dini penyakit diabetes mayoritas teknik yang digunakan adalah klasifikasi. Teknik klasifikasi yaitu mengenali pola atau model dari sebuah dataset khususnya dataset penyakit diabetes. Tujuannya agar model tersebut dapat digunakan untuk memprediksi ataupun klasifikasi apakah seseorang tersebut menderita penyakit jantung atau tidak. Model tersebut didasarkan pada analisis data training. Model dari hasil klasifikasi dapat dimanfaatkan untuk mengklasifikasikan serta memprediksi tren data masa depan [7].

Terdapat penelitian yang melakukan tinjauan komprehensif tentang teknik data mining khususnya dalam diagnosa dan prediksi diabetes berdasarkan teknik klasifikasi yang umum digunakan [8]. Dalam penelitian tersebut dilakukan evaluasi berdasarkan beberapa parameter seperti algoritma/model, tipe data, kemampuan plug-n-play, dll. Berdasarkan evaluasi tersebut, dapat disimpulkan bahwa agar menghasilkan deteksi, klasifikasi dan prediksi penyakit yang akurat perlu dilakukan preprocessingi data dan menggunakan teknik hybrid dimana menggabungkan model yang berbeda secara parallel daripada menggunakan model individu. Oleh karena itu pada penelitian ini akan menggunakan metode ensemble atau meta learning. Metode ensemble ialah gabungan dari beberapa algoritma/model yang dijadikan satu sehingga kinerjanya lebih baik dibandingkan single classifier. Algoritme ensemble yang digunakan pada penelitian ini adalah Stacking, dengan tujuan untuk meningkatkan kinerja single classifier. Dengan adanya peningkatan kinerja klasifikasi diharapkan dapat menghasilkan akurasi yang optimal, sehingga dapat digunakan sebagai referensi dalam melakukan penelitian, terutama dalam penelitian yang mengembangkan berbagai sistem yang dapat meningkatkan keberhasilan deteksi dini penyakit diabetes.

2. TINJAUAN PUSTAKA

2.1. Diabetes Melitus

Penyakit diabetes mellitus dibagi menjadi 2 jenis, yaitu Diabetes mellitus Tipe 1 (tipe A) dan Diabetes mellitus Tipe 2 (tipe B). Penyakit diabetes mellitus tipe 1 biasa disebut insulin dependent. Diabetes mellitus tipe 1 ini biasa terjadi pada usia muda dibawah 30 tahun. Seseorang yang menderita diabetes mellitus tipe ini memerlukan tindakan suntik insulin. Suntik insulin dilakukan karena glukosa darah dalam tubuh tidak dapat memproduksi insulin sebagaimana mestinya [9]. Sedangkan diabetes melitus tipe 2 ini biasa disebut non-insulin dependent yang ditandai dengan resistensi insulin dan gangguan sekresi insulin. Tipe ini seringkali terjadi pada penderita yang berusia diatas 40 tahun. Hal ini terjadi ketika tubuh tidak bisa lagi secara aktif menggunakan insulin yang dihasilkan oleh tubuh. Biasanya disebabkan faktor keturunan, obesitas, kurang aktivitas, penyakit lain dan usia [10].

2.2. Faktor Penyakit Diabetes Melitus

Terdapat beberapa faktor risiko pada penyakit diabetes mellitus seperti [11]: Nafsu Makan Meningkat, Sering Buang Air Kecil, Peningkatan Kehausan, Turunnya Berat Badan, Usia (15-40) tahun, Faktor Keturunan, Mulut Kering, Mudah Kelelahan/Kurangnya Aktivitas Fisik, Sering Mengantuk, Mual/Muntah-Muntah, Timbulnya Luka yang Tak Kunjung Sembuh, Gatal-Gatal, Mengonsumsi makanan Berkolesterol Tinggi, Obesitas, Kadar Glukosa Darah Meningkat.

2.3. Data Mining

Metode data mining merupakan sebuah proses menentukan ikatan yang mengandung arti, pola, dan keterkaitan dengan mengolah kelompok data. Dalam data mining terdapat 6 metode yang biasa di jalankan yaitu ramalan atau prediksi, penggambaran atau deskripsi, klasifikasi, estimasi, asosiasi dan clustering [12].

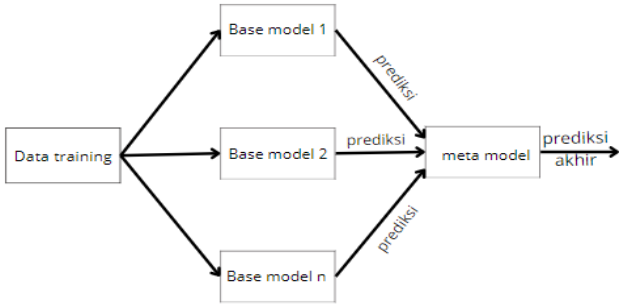
Penggunaan data mining memiliki dampak diberbagai bidang seperti bidang kesehatan. Data mining pada system medis dibutuhkan untuk mengekstrak informasi dari data base sehingga dapat melakukan diagnosis penyaikit [13].

2.4. Metode Klasifikasi

Klasifikasi atau pengelompokan adalah teknik pada data mining untuk mengenali suatu karakteristik dari data kemudian data tersebut dikelompokan dalam beberapa kelas yang sudah didefinisikan [12]. Metode klasifikasi banyak digunakan untuk berbagai aplikasi, diantaranya untuk mendeteksi suatu penyakit, pengelolaan barang, prediksi penjualan dan lain sebagainya.

2.5. Metode *Stacking*

stacking adalah sebuah metode dimana sebuah learner ditraining untuk menggabungkan beberapa individual learner yang disebut first-level learner, sedangkan yang digabungkan disebut second-level learner atau meta-learner [14]. Stacking melibatkan dua atau lebih base learner sebagai model level-0, dan meta-learner yang menggabungkan prediksi base learner disebut sebagai level-1. Model yang berbeda digunakan oleh base learner untuk belajar dari suatu dataset. Output dari masing-masing model dikumpulkan untuk membuat dataset baru. Di dalam dataset yang baru, setiap sampel berhubungan dengan nilai sesungguhnya yang seharusnya diprediksi. Selanjutnya dataset tersebut digunakan oleh stacking model learner level-1 hingga mendapatkan hasil akhir.



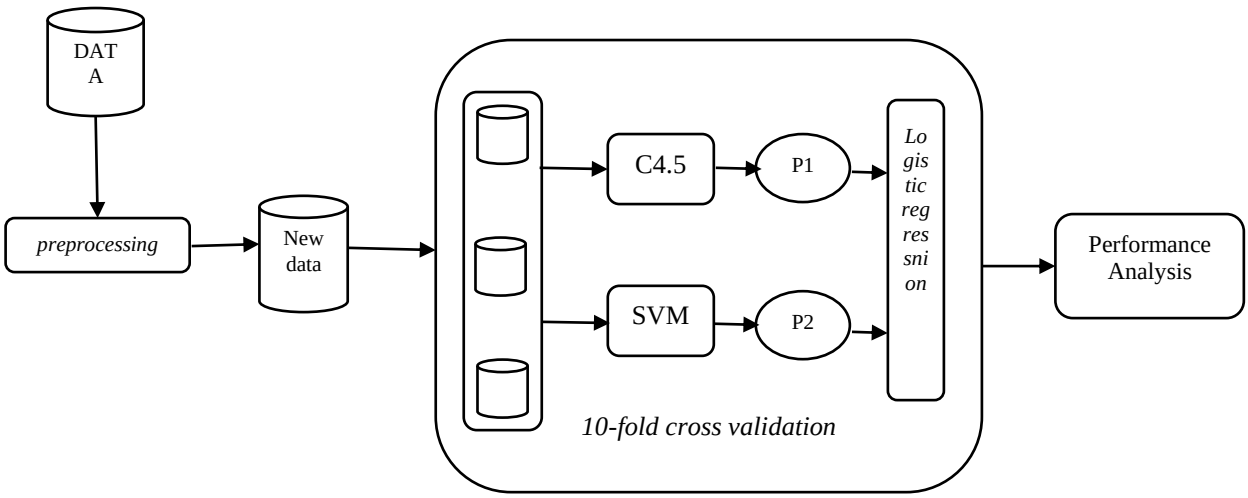
Gambar 1. Tahapan Metode *Stacking*

Langkah-langkah pada metode stacking yaitu:

1. Memisahkan data menjadi dua bagian
2. Melakukan training pada beberapa base-learner menggunakan bagian data 1
3. Membuat prediksi dengan base-learner menggunakan data 2
4. Menggunakan hasil prediksi dari Langkah 3 sebagai input (data train) learner kedua.

3. METODOLOGI PENELITIAN

Penelitian ini akan dilakukan berdasarkan tahapan pengumpulan data, preprocessing data, klasifikasi, hingga evaluasi sebagaimana ditunjukan pada gambar berikut.



Gambar 2. Metodologi Penelitian

Tahap pertama adalah menyiapkan dataset, kemudian dilakukan preprocessing data. Selanjutnya data baru yang telah dilakukan preprocessing dilakukan proses pembentukan model klasifikasi. Model yang digunakan yaitu metode stacking dan validasi menggunakan 10 fold cross validation. Selanjutnya dievaluasi menggunakan confussion matrix dan menghasilkan akurasi.

3.1. Sumber Data

Pada penelitian ini proses pertama adalah pengumpulan data. Dataset yang digunakan merupakan dataset public bersumber dari repository UCI Machine Learning yaitu Early stage diabetes risk prediction dataset yang dapat diakses di https://archive.ics.uci.edu/ml/machine-learning-databases/00529/. Dataset ini terdiri dari 17 variabel dengan jumlah data sebanyak 520. Dataset ini dikumpulkan berdasarkan kuesioner yang dibagikan langsung kepada pasien dari Sylhet Diabetes Hospital of Sylhet, Bangladesh. Para pasien ini termasuk orang-orang yang baru saja menjadi penderita diabetes, atau yang memiliki gejala merujuk ke diabetes. Berikut ini adalah tabel yang merupakan detail variabel pada dataset yang digunakan dalam penelitian ini.

Tabel 1. Variabel Data Penelitian

No.	Atribut	Tipe
1	Umur	Numeric
2	Jenis kelamin	Catergorical
3	*Polyuria*	Catergorical
4	*Polydipsia*	Catergorical
5	*Suddenweight loss*	Catergorical
6	*Weakness*	Catergorical
7	*Polyphagia*	Catergorical
8	*Genital thrush*	Catergorical
9	*Visual blurring*	Catergorical
10	*Itching*	Catergorical
11	*Irritability*	Catergorical
12	*Delayed healing*	Catergorical
13	*Partial paresis*	Catergorical
14	*Muscle stiffness*	Catergorical
15	*Alopecia*	Catergorical
16	Obesitas	Catergorical
17	Kelas	Catergorical

3.2. *Preprocessing Data*

Tahap selanjutnya setelah pengumpulan data adalah dilakukan preprocessing data. Tahap preprocessing data yaitu meliputi pengisian data yang kosong, menghapus duplikasi data, dan memeriksa inkonsistensi data.

3.3. Proses Klasifikasi

Setelah data melalui proses preprocessing maka dilanjutkan dengan proses pemodelan klasifikasi. Dataset akan diklasifikasi menggunakan masing-masing algoritma klasifikasi C4.5 dan Support Vector Machice (SVM) secara individu juga menggunakan metode ensemble yaitu stacking dimana menggabungkan antara algoritma C4.5 dan SVM sebagai base model kemudian hasil dari masing-masing algoritma akan dijadikan input metode selanjutnya yang menjadi meta model dalam hal ini adalah Logistic Regression. Teknik cross validation digunakan untuk memvalidasi keakuratan sebuah model, dalam penelitian ini menggunakan teknik 10-fold cross validation dimana data dibagi menjadi 10 bagian yang memiliki rasio yang hampir sama.

3.4. *Performance Measure*

Confusion matrix adalah tabel yang berisikan informasi mengenai perbandingan hasil klasifikasi antara aktual dan prediksi untuk setiap model algoritma klasifikasi.

Tabel 2. Confusion Matrix

Actual	**Prediksi**	
	True	**False**
True	TP	FN
False	FP	TN

Indikator evaluasi yang digunakan untuk mengukur performa dalam penelitian ini yaitu:

1. *Accuracy*, yaitu presentase dari jumlah prediksi yang benar dari seluruh jumlah prediksi yang dihasilkan oleh classifier.

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100\%$$

2. *Recall*, yaitu presentase dari prediksi *true positive*, dibandingkan dengan keseluruhan data *positive*.

$$recall = \frac{TP}{TP + FN} \times 100\%$$

3. *Precision*, yaitu ukuran presentase dari prediksi *true positive* dibandingkan seluruh hasil yang diprediksi *positive.*

$$precision = \frac{TP}{TP + FP} \times 100\%$$

4. HASIL DAN PEMBAHASAN

Dalam penelitian ini telah dilakukan proses klasifikasi/prediksi penderita penyakit diabetes dengan menggunakan metode Stacking dengan algoritma C4.5, SVM sebagai base model dan algoritma logistic regression sebagai meta model. Berdasarkan proses klasifikasi tersebut performanya akan dibandinglan dengan hasil yang didapatkan oleh base/single classifier. Berikut tabel yang menjelaskan hasil klasifikasi berupa tabel confunsion matrix masing-masing algoritma.

Tabel 3. Confusion Matrix masing-masing Algoritma

	TP	**TN**	**FP**	**FN**
C4.5	304	191	16	9
SVM	309	193	11	7
Stacking	312	193	8	7

Kemudian dari tabel confusion matrix tersebut dapat dilakukan evaluasi dengan menghitung nilai accuracy, recall, dan precision pada masing-masing algoritma. Di bawah ini tabel dari hasil perhitungan masing-masing algoritma.

Tabel 4. Hasil Perhitungan *accuracy, recall* dan *precision* masing-masing Algoritma

	C4.5	**SVM**	***Stacking***
Accuracy	95,19 %	96, 53 %	97, 11 %
Recall	97, 12 %	97,78 %	97, 80 %
Precision	95, 00 %	96, 56 %	97, 50 %

Tabel di atas menunjukan berdasarkan pengujian bahwa metode atau algoritma yang diusulkan yaitu metode Stacking berhasil meningkatkan nilai accuracy, recall dan precision dibandingkan dengan base/single classifier.

5. KESIMPULAN DAN SARAN

Berdasarkan penelitian yang telah dilakukan dengan metode Stacking menggunakan C4.5 dan SVM sebagai base model dengan logistic regression sebagai meta model dalam klasifikasi penderita penyakit diabetes, menunjukan bahwa algoritma Stacking mampu meningkatkan performa base/single classifier dari sisi accuracy, recall dan precision.

Dengan adanya peningkatan nilai tersebut diharapkan penelitian ini dapat menjadi referensi untuk penelitian atau pengembangan system yang dapat memaksimalkan tingkat keberhasilan deteksi dini penderita penyakit diabetes. Penggunaan metode stacking yang lebih baik dapat dilakukan pada penelitian selanjutnya untuk meningkatkan performa klasifikasi pada penderita diabetes dengan memodifikasi pada base model dan meta model nya.

DAFTAR PUSTAKA

[1] Khairani, "Pengetahuan Diabetes Mellitus Dan Upaya Pencegahan Pada Lansia Di Lam Bheu Aceh Besar," Pengetah. Diabetes Mellit. Dan Upaya Pencegah. Pada Lansia Di Lam Bheu Aceh Besar, vol. 3, no. 3, pp. 58– 66, 2012.
[2] Nurlina, "Jurnal Media Keperawatan : Politeknik Kesehatan Makassar Jurnal Media Keperawatan : Politeknik Kesehatan Makassar," J. Media Keperawatan Politek. Kesehat. Makassar, vol. 10, no. 01, pp. 59– 66, 2019.
[3] Ulfa, N. M., Lubada, E. I. and Darmawan, R. (2020) Buku Ajar Farmasi Klinis dan Komunitas : Medication Picture dan Pill Count Pada Kepatuhan Minum Obat Penderita Diabetes Mellitus dan Hipertensi. 1st edn. Gresik: Graniti.
[4] ADA, "Classification and Diagnosis of Diabetes Mellitus," Stand. Med. Care Diabetes, vol. 39, no. January, 2016, doi: 10.1016/B978-0-323-18907-1.00038-X.

[5] "Retinopati Diabetik" [Daring]. Tersedia pada: https://www.alodokter.com/retinopati- diabetik. [Diakses: 4-Juni-2020].
[6] S. J. Russell, P. Norvig, J. F. Canny, J. M. Malik, and D. D. Edwards, "Artificial Intelligence: A Modern Approach", vol. 2. Prentice hall Englewood Cliffs, 1995.
[7] I. H. Witten, E. Frank, and M. A. Hall, Data Mining: Practical Machine Learning Tools and Techniques, 3rd ed. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2011.
[8] F. A. Khan dkk, Detection and Prediction of Diabetes Using Data Mining: A Comprehensive Review
[9] Sa'di,S.,Maleki, A., Hashemi,R., Panbechi,Z., & Chalabi, K.(2015). Comparison of Data Mining Algorithmsin the Diagnosisof TypeIiDiabetes. InInternational Journal on Computational Science & Applications (Vol. 5,hal. 1–12).
[10] V.,&Ravikumar,A.(2014). Study of Data Mining Algorithms for Prediction and Diagnosis of Diabetes Mellitus. International Journal of Computer Application,95(17), 12–16.
[11] Devi,M.R.,& Shyla,J.M. (2016). Analysisof Various Data Mining Techniques to Predict Diabetes Mellitus. International Journal of Applied Engineering Research, 11, 727–730.
[12] Larose, D. T. and Larose, C. D. (2014) Discovering Knowledge in Data: An Introduction to Data Mining: Second Edition, Discovering Knowledge in Data: An Introduction to Data Mining: Second Edition. doi: 10.1002/9781118874059.
[13]. Poonguzhali,E., Kabilan,S.,Kannan,S., & Sivagami, P.(2014). Diagnosisof Diabetes Mellitus Type2using NeuralNetwork. International Journal of Emerging Technology and Advanced Engineering,4(2),939–942.
[14] Z.H. Zhou, Ensemble Methods Foundations and Algorithms. CRC Press, 2012.