

Perbandingan Kinerja Rule ZeroR Dan Function SMO Dengan T-Test Dalam Pengklasifikasian Diagnosis Penyakit Diabetes Mellitus

Slamet Wiyono¹

¹ DIV Teknik Informatika Politeknik Harapan Bersama
slamet2wiyono@gmail.com

ABSTRAK

Penemuan informasi dari data medis adalah salah satu cara untuk membuat keputusan dari pasien.. Dataset diabetes mellitus diperoleh dari Pima Indian dataset diabetes dari repositori UCI. Pengolahan data mining dibagi menjadi dua tahap yaitu pertama, identifikasi dan pencarian atribut data , pencarian keakurasian data menggunakan software WEKA dan kedua, tahap kelayakan perbandingan kedua data tersebut dengan t-test menggunakan Microsoft excell. Masing-masing metode diuji keefektifitasnya dengan menggunakan 10-fold cross validation dengan hasil pengukuran diperoleh akurasi 77,3 % untuk function SMO dan 65,1 % untuk rule ZeroR . Setelah didapatkan hasil akurasi, dilakukan pengujian perbandingan kesignifikaan dengan confidence level 5% dengan menggunakan t-test. Didapatkan hasil bahwa keduanya signifikan, hal ni berarti bahwa metode function SMO memberikan prediksi yang lebih baik dari rule ZeroR untuk diagnose penyakit diabete.

Kata Kunci : *t-test, function SMO, ZeroR.*

1. PENDAHULUAN

Terdapat banyak sekali metode yang terdapat dalam teknik learning dengan berbagai variasinya yang telah diusulkan dan diimplementasikan, diantaranya *rule ZeroR* dan *function SMO*. DM merupakan penyakit yang terjadi akibat kadar glukosa di dalam darah tinggi karena tubuh tidak dapat melepaskan atau menggunakan insulin secara normal. Kadar glukosa darah sepanjang hari bervariasi, meningkat setelah makan dan kembali normal dalam waktu dua jam. Glukosa darah normal pada pagi hari setelah malam sebelumnya berpuasa adalah 70-110 mg/dL. Glukosa darah biasanya kurang dari 120-140 mg/dL pada dua jam setelah makan atau minum cairan yang mengandung gula maupun karbohidrat lainnya. Glukosa darah normal cenderung meningkat secara ringan tetapi progresif setelah usia 50 tahun, terutama pada orang – orang yang tidak aktif beraktifitas. Insulin adalah hormon yang dilepaskan oleh pankreas, merupakan zat utama yang bertanggungjawab dalam mempertahankan kadar glukosa darah yang tepat. Insulin menyebabkan glukosa berpindah ke dalam sel sehingga bisa menghasilkan energi. DM terjadi

jika tubuh tidak menghasilkan insulin yang cukup untuk mempertahankan glukosa darah normal atau jika sel tidak memberikan respon yang tepat terhadap insulin.

Terdapat dua tipe DM, yaitu DM tipe 1 yang merupakan diabetes yang tergantung pada insulin, dimana pankreas menghasilkan sedikit insulin atau sama sekali tidak menghasilkan insulin. Sedangkan pada DM tipe 2, pankreas tetap menghasilkan insulin tetapi kadang kadarnya lebih tinggi dari normal dimana kejadian ini akan menyebabkan tubuh membentuk kekebalan terhadap efeknya, sehingga kekurangan insulin relatif. Gejala awal dari DM ini biasanya diawali oleh tiga kondisi, yaitu poliuri (meningkatnya pengeluaran kemih), polidipsi (rasa haus yang berlebihan), dan polifagi (meningkatnya rasa lapar).

Banyak penyandang DM yang terdiagnosis setelah mengalami komplikasi. Padahal, apabila dilakukan diagnosis secara dini, maka penanganan bisa dilakukan lebih cepat dan komplikasi yang membahayakan dapat dihindari. Dalam perkembangan di dunia kedokteran saat ini, para peneliti dan praktisi memusatkan perhatiannya untuk mendeteksi kondisi DM dan mencegah atau menghambat

berkembangnya komplikasi. Untuk mendukung hal ini dapat digunakan teknik data mining untuk menggali informasi yang berharga dari kumpulan informasi diabetes. Dalam penelitian ini dilakukan data mining dari dataset DM kelompok suku Pima Indians, Amerika Serikat, dimana berdasarkan penelitian yang dilakukan oleh National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) sejak tahun 1965 lebih dari 50% populasinya menderita diabetes tipe 2 dan rata – rata angka kematian akibat DM ini 10 kali lebih besar dibandingkan populasi lainnya di Amerika Serikat [1, 2]. Dataset Pima meliputi sembilan atribut pengukuran dari pasien dengan DM tipe 2 positif dan pasien dengan diagnosis DM negatif. Dalam penelitian ini akan dibandingkan sistem pendukung keputusan untuk diagnosis DM menggunakan decision tree ID3 dan J48 dengan melakukan ekstraksi informasi dalam bentuk pohon telusur (tree) dari dataset Pima. Untuk mengetahui efektifitas dan akurasi dari pengklasifikasi decision tree ini maka dilakukan analisa dari matrik confusion (Bradski dan Kachler, 2008). [3]

2. METODE PENELITIAN

Penelitian ini terbagi menjadi dua tahap yaitu pertama, identifikasi dan pencarian atribut data , pencarian keakurasian data menggunakan software WEKA dan kedua, tahap kelayakan perbandingan kedua data tersebut dengan t-test menggunakan Microsoft excell. Data yang digunakan dalam penelitian ini yaitu data bebas yang berasal dari Pima Indians Diabetes Dataset, UCI Machine Learning Repository.

3. PROSES

A. Identifikasi dan Pemilihan Atribut

Dataset dalam penelitian ini diambil dari repositori database Pima Indians, UCI. Tabel 1 menjelaskan atribut dataset diabetes Pima Indians. Dataset Pima ini terdiri dari 768 data klinis yang semuanya berasal dari jenis kelamin wanita dengan umur sekurang – kurangnya 21 tahun. Penggunaan setiap atribut pada dataset

Pima ini akan memberikan hasil yang berbeda – beda pada akurasi diagnosis DM dan hal ini berkaitan ada atau tidaknya kelengkapan nilai dari setiap atribut.

Tabel.1 Atribut dataset diabetes Pima Indians

<i>Singkatan</i>	<i>Deskripsi</i>	<i>Satuan</i>
Pregnant Glucose	Banyaknya kehamilan Kadar glukosa dua jam setelah makan	- Mg/dL
DBP	Tekanan darah	Mm Hg
TSFT	Ketebalan kulit	mm
INS	Insulin	mu U/ml
BMI	Berat Tubuh	Kg/m ²
DPF	Riwayat diabetes dalam keluarga	-
Age	Umur	Years
Class	Positif diabetes (1) dan negatif diabetes (0)	-

B. Pengolahan Data

Hasil pengolahan dan uji coba menggunakan rule ZeroR dan function SMO pada dataset dihasilkan informasi seperti pada Tabel 2

Tabel 2. Hasil pengujian zeroR dan SMO dengan 10-fold cross validation

	Correct	Incorrect	Percent correct
Rule ZeroR	500	268	65,1%
Function SMO	594	174	77,3%

Untuk mengetahui perbandingan kedua pengujian tersebut signifikan atau tidak maka dilakukan perhitungan dengan t-test dengan confidence level 5%. Hasil t-test ditunjukkan pada Gambar.1

t-Test: Two-Sample Assuming Equal Variances		
	Variable 1	Variable 2
Mean	77.3445	65.10595
Variance	16.52427	0.12978
Observati	10	10
Pooled Va	8.327023	
Hypothesi	0	
df	18	
t Stat	9.483532	
P(T<=t) or	1E-08	
t Critical o	1.734064	
P(T<=t) tw	2.01E-08	
t Critical t	2.100922	

Gambar.1 hasil perhitungan t-test menggunakan t-test.

4 KESIMPULAN

Perbandingan kinerja rule ZeroR dan fuction SMO dilakukan dengan menggunakan software WEKA kemudian dicari perbedaanya menggunakan t-test. SMO memiliki keakurasian 77,3% dan ZeroR sebesar 65.1% sehingga didapatkan hasil bahwa kinerja function SMO lebih baik dari ZeroR.

DAFTAR PUSTAKA

[1] National Institute of Diabetes and Digeative and Kidney Diseases, “Conquering Diabetes, A Strategic Plan for the 21st Century,” National Institutes of Health, U.S. Department of Health and Human Services, NIH Publication No. 99-4398, 1999.

[2] Smith, J.W., Everhart, J.E., Dickson, W.C., Knowler, W.C., & Johannes, R.S., “Using the ADAP learning algorithm to forecast the onset of diabetes mellitus,” in *Proceedings of the Symposium on Computer Applications and Medical Care*, IEEE Computer Society Press, pp. 261-265, 1988.

[3] Zhu, W., Zeng, N., & Wang, N., “Sensitivity, specificity, accuracy confidence interval and ROC analysis with practical SAS implementations,” Nesug, Health Care and Life Sciences, pp. 1-9, 2010.

[4] WEKA, Machine Learning Group at University of Waikato, diambil dari <http://www.cs.waikato.ac.nz/ml/weka/>

[5] Pima Indians Diabetes Dataset, UCI Machine Learning Repository, diambil dari <http://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes>