

PERBANDINGAN METODE NEURAL NETWORK DAN SUPPORT VECTOR MACHINE DALAM KLASIFIKASI DIAGNOSA PENYAKIT DIABETES

Dhita Diana Dewi¹, Nurul Qisthi², Siti Sarah Sobariah Lestari³,
Zulfa Hidayah Satria Putri⁴

^{1,3} Universitas Nusa Putra

² STIKom Poltek Cirebon

⁴ Badan Pusat Statistik RI

E-mail : dhita.dianadewi@nusaputra.ac.id, nurulqisthi@stikompoltek.ac.id,
siti.sobariah@nusaputra.ac.id, zulfahsp@bps.go.id

Kata Kunci

Neural Network, Support Vector Machine, Diabetes.

Abstrak

Terdapat banyak penelitian yang telah dilakukan untuk memprediksi atau mendiagnosa awal penyakit sesuai dengan kondisi yang berkaitan. Penelitian tersebut dilakukan dengan teknik data mining yang terdiri dari berbagai algoritma seperti Naive Bayes, Neural Network, Support Vector Machine (SVM) dan lain-lain. Dengan adanya diagnosis awal atau prediksi, diharapkan dapat menghindari atau mencegah hal yang membahayakan, pada kasus ini adalah terkena penyakit diabetes. Penelitian ini bertujuan untuk mengetahui pengklasifikasian diagnosis penyakit diabetes dan membandingkan dua algoritma yang digunakan yaitu Neural Network dan Support Vector Machine untuk mendapatkan algoritma dengan tingkat akurasi paling baik dalam pengklasifikasian diagnosis penyakit diabetes. Data yang digunakan dalam penelitian ini adalah data sekunder. Data tersebut berjumlah 768 observasi dengan menampilkan 9 atribut, yaitu Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, BMI, Diabetes Pedigree Function, Age dan Outcome. Teknik klasifikasi data mining yang digunakan dalam penelitian ini adalah Neural Network dan Support Vector Machine (SVM). Evaluasi dan validasi model pada penelitian ini menggunakan metode K-Fold Cross Validation (K-Fold=10). Berdasarkan hasil analisis pada penelitian ini didapatkan nilai akurasi untuk metode ANN sebesar 77,60%, sedangkan nilai akurasi untuk metode SVM sebesar 65,24%. Artinya penggunaan metode ANN lebih baik daripada SVM untuk mengklasifikasikan seseorang menderita diabetes atau tidak. Sedangkan untuk metode ANN memiliki nilai AUC sebesar 0,834 sehingga dapat dikategorikan dalam good classification.

Keywords

Neural Network, Support Vector Machine, Diabetes.

Abstract

There have been many studies conducted to predict or diagnose early diseases based on relevant conditions. These studies employ data mining techniques involving various algorithms such as Naive Bayes, Neural Network, Support Vector Machine (SVM), and others. With early

diagnosis or prediction, it is hoped to avoid or prevent harmful occurrences, in this case, the onset of diabetes. This research aims to understand the classification of diabetes disease diagnosis and compare two algorithms used, namely Neural Network and Support Vector Machine, to determine the algorithm with the highest accuracy rate in classifying diabetes disease diagnosis. The data used in this study is secondary data. The data consists of 768 observations with 9 attributes: Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, BMI, Diabetes Pedigree Function, Age, and Outcome. The data mining classification techniques employed in this study are Neural Network and Support Vector Machine (SVM). Model evaluation and validation in this study are performed using the K-Fold Cross Validation method (K-Fold=10). Based on the analysis results of this study, the accuracy value for the ANN method is 77.60%, while the accuracy value for the SVM method is 65.24%. This indicates that the use of the ANN method is superior to SVM in classifying whether someone has diabetes or not. Additionally, the ANN method has an AUC value of 0.834, categorizing it as a good classification.

*Correspondent Author: Dhita Diana Dewi
Email : dhita.dianadewi@nusaputra.ac.id



PENDAHULUAN

Pada setiap tahun di seluruh dunia, jutaan manusia meninggal karena terkena atau mengidap beragam penyakit, salah satunya adalah karena penyakit diabetes mellitus. Menurut data World Health Organization (WHO) pada tahun 2016, penyakit diabetes merupakan salah satu diantara 10 penyakit mematikan atau yang menyebabkan kematian terbanyak di seluruh dunia. Berdasarkan data WHO tahun 2015, penyakit ini menyumbangkan 2,8 % atau 1,6 juta dari jumlah kematian penduduk dunia secara keseluruhan. Jumlah mengalami peningkatan, yaitu dari 1 juta jiwa pada tahun 2000. Kondisi tersebut juga dialami di Indonesia. Ancaman penyakit diabetes mellitus sangatlah berbahaya. Berdasarkan data Riset Kesehatan Dasar (Riskesdas) 2018, tingkat prevalensi penyakit ini naik dari 6,9% pada tahun 2013 menjadi 8,5% pada tahun 2018. Kondisi ini mengakibatkan estimasi harapan hidup berkurang sekitar 5 sampai 10 tahun.

Menurut World Health Organization (WHO), Diabetes Mellitus adalah suatu penyakit kronis dimana organ pankreas tidak memproduksi cukup insulin atau ketika tubuh tidak efektif dalam menggunakannya. Definisi diabetes lainnya menurut ahli, yaitu Diabetes mellitus merupakan penyakit metabolik yang berlangsung lama atau kronis yang ditandai dengan peningkatan kadar gula darah sebagai akibat dari kelainan insulin, aktivitas insulin ataupun sekresi insulin yang dapat menimbulkan berbagai masalah serius dan prevalensi dari penyakit diabetes mellitus ini berkembang sangat cepat (Smeltzer
*PERBANDINGAN METODE NEURAL NETWORK DAN SUPPORT VECTOR MACHINE
DALAM KLASIFIKASI DIAGNOSA PENYAKIT DIABETES*

&Bare, 2008). Terdapat empat kelas klinis klasifikasi diabetes mellitus menurut penyebab atau etiologi (Perkeni, 2011) adalah sebagai berikut.

Tabel 1. Klasifikasi Diabetes Mellitus

Jenis	Penyebab
Tipe 1	Kerusakan sel beta pankreas, umumnya mengarah ke defisiensi insulin absolut, biasanya disebabkan oleh autoimun dan idiopatik.
Tipe 2	Penyebab bisa bervariasi, bisa disebabkan oleh resistensi insulin yang disertai insulin relatif sampai dengan defek sekresi insulin disertai resistensi insulin.
Tipe lain	Bisa disebabkan oleh defek genetik fungsi beta, defek genetik kerja insulin, penyakit endokrin pankreas, oleh karena obat-obatan, infeksi, ataupun penyakit genetik lainnya.
Diabetes mellitus gestasional	Intoleransi glukosa yang timbul atau terdeteksi pada kehamilan pertama dan gangguan toleransi glukosa setelah terminasi kehamilan.

Hal yang perlu menjadi perhatian adalah diabetes dapat menyebabkan komplikasi yang mampu berdampak buruk bagi fungsi mata, jantung, ginjal, kulit, saraf, sampai saluran pernapasan. Diagnosis awal dapat menjadi Langkah yang tepat untuk menghindari penyakit ini. Menurut American Diabetes Association (ADA, 2015) menyatakan bahwa diagnosis DM (Diabetes Mellitus) dapat dilakukan dengan melihat manifestasi berupa gejala DM (poliuria, polidipsia, polifagia, penurunan berat badan tanpa sebab) ditambah dengan kadar glukosa darah sewaktu >200 mg/dL, atau kadar glukosa darah puasa >126 mg/dL atau kadar glukosa darah 2 jam setelah dilakukan test toleransi glukosa oral (75 gram glukosa yang dilarutkan) makan >200 mg/dL. Pemeriksaan dilakukan minimal 2 kali dengan cara yang sama.

Selain itu, terdapat banyak penelitian yang telah dilakukan untuk memprediksi atau mendiagnosa awal penyakit sesuai dengan kondisi yang berkaitan. Penelitian tersebut dilakukan dengan teknik data mining yang terdiri dari berbagai algoritma seperti Naive Bayes, *Neural Network*, *Support Vector Machine* (SVM) dan lain-lain. Dengan adanya diagnosis awal atau prediksi, diharapkan dapat menghindari atau mencegah hal yang membahayakan, pada kasus ini adalah terkena penyakit diabetes. Oleh karena itu, penelitian ini bertujuan untuk mengetahui pengklasifikasian diagnosis penyakit diabetes dan membandingkan dua algoritma yang digunakan yaitu *Neural Network* dan *Support Vector Machine* untuk mendapatkan algoritma dengan tingkat akurasi paling baik dalam pengklasifikasian diagnosis penyakit diabetes.

Tujuan dari penelitian ini adalah sebagai berikut : 1. Membandingkan dua algoritma yang digunakan yaitu *Neural Network* dan *Support Vector Machine* untuk mendapatkan algoritma dengan tingkat akurasi paling baik dalam pengklasifikasian diagnosis penyakit diabetes, dan Mengetahui pengklasifikasian diagnosis penyakit diabetes.

Hasil penelitian ini yang berupa klasifikasi diagnosis penyakit diabetes dapat bermanfaat dan digunakan sebagai suatu faktor penanda/ciri atau prediksi diagnosis awal penyakit diabetes yang diharapkan dapat menjadi dasar atau langkah awal melakukan upaya pencegahan penyakit diabetes.

METODE PENELITIAN

Pada penelitian ini, data yang digunakan adalah berupa data sekunder yang berasal dari database kesehatan Pima Indians Diabetes Dataset yang diambil dari database *website* <https://www.kaggle.com/uciml/pima-indians-diabetes-database>. Data tersebut berjumlah 768 observasi dengan menampilkan 9 atribut, yaitu Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, BMI, Diabetes Pedigree Function, Age dan Outcome. Definisi variabel atau atribut yang digunakan pada penelitian adalah sebagai berikut :

Tabel 1. Definisi Operasional Varibel/Atribut

No	Nama Variabel/Atribut (label)	Definisi	Tipe
1	Pregnancies	Banyaknya kehamilan yang pernah dialami	Numerik
2	Glucose	Konsentrasi plasma glukosa dalam 2 jam tes toleransi oral glukosa (mg/dL)	Numerik
3	Blood Pressure	Tekanan darah diastolik (mm Hg)	Numerik
4	Skin Thickness	Ketebalan lipatan kulit trisep (mm)	Numerik
5	Insulin	Insulin serum 2 Jam (mu U/ml)	Numerik
6	BMI	<i>Body Mash Index</i> (BMI) atau Indeks massa tubuh. $BMI = \frac{\text{Berat Badan (Kg)}}{(\text{Tinggi badan (m)})^2}$	Numerik
7	Diabetes Pedigree Function	Fungsi silsilah diabetes	Numerik
8	Age	Umur tunggal (Tahun)	Numerik
9	Outcome	Status penderita penyakit Diabetes Mellitus (Yes/No) yang merupakan Variabel kelas	Kategorik Tranformasi menjadi Numerik (0 dan 1) : 1 = penderita

	Penyakit DM (268 Observasi)
	0 = bukan penderita
	Penyakit DM (500 observasi)

Proses mengubah kategori data dari data kategori berupa karakter menjadi numerik (angka) dilakukan sebagai persiapan untuk memudahkan pengolahan pembuatan model ANN dan SVM. Pengolahan dan analisis data pada penelitian ini menggunakan *Software RapidMiner Studio*.

Dalam data mining, metode yang dapat digunakan untuk memprediksi berdasarkan data masa lalu adalah klasifikasi (Larose, 2005). Klasifikasi merupakan operasi untuk memisahkan beragam entitas kedalam beberapa kelas (Nisbet, 2009: 236). Terdapat beberapa Teknik klasifikasi dalam data mining, namun yang akan digunakan dalam penelitian ini adalah *Neural Network* dan *Support Vector Machine (SVM)*.

2.1 Artificial Neural Network (ANN)

Artificial Neural Network (ANN) atau Jaringan Saraf Tiruan (JST) menurut Haykin (2009), adalah prosesor yang terdistribusi besar – besaran secara parallel yang dibuat dari unit proses sederhana, yang mempunyai kemampuan untuk menyimpan pengetahuan berupa pengalaman dan

apat digunakan untuk proses lain. Selain itu, ANN juga dapat didefinisikan sebagai sebuah jaringan yang dirancang untuk menyerupai otak manusia yang bertujuan untuk melaksanakan suatu tugas tertentu. ANN dapat digunakan untuk memodelkan hubungan yang kompleks antara input dan output untuk menemukan pola-pola pada data. Jaringan ini biasanya diimplementasikan dengan menggunakan komponen elektronik atau disimulasikan pada aplikasi komputer.

Menurut Fausett (1994), ANN dikembangkan sebagai generalisasi model matematika yang meniru jaringan saraf biologi memiliki asumsi, yaitu sebagai berikut.

1. Proses informasi yang terjadi di banyak elemen sederhana yang disebut neuron.
2. Sinyal dikirimkan melalui hubungan koneksi di antara neuron – neuron.
3. Setiap hubungan koneksi mempunyai bobot yang saling berhubungan, di mana jaringan saraf yang khas memperbanyak sinyal yang ditransmisi.
4. Setiap neuron mengaplikasikan sebuah fungsi aktivasi (biasanya berupa non – linear) ke jaringan inputnya (jumlah dari bobot sinyal input) untuk menentukan sinyal outputnya.

Tahapan proses alur kerja algoritma *backpropagation* ANN adalah sebagai berikut:

1. Inisialisasi semua bobot awal jaringan. Inisialisasi bobot jaringan biasanya menggunakan bilangan acak dengan jangkauan [-0.5,0.5] atau menggunakan distribusi uniform dalam jangkauan kecil.

- Menjumlahkan sinyal-sinyal input terbobot, yaitu setiap data pada data training, hitung input untuk simpul berdasarkan nilai input dan bobot jaringan saat itu, menggunakan rumus:

$$s_j = \sum_{i=1}^n w_{ij} x_i + b_j \quad (1)$$

- Selanjutnya, hasil input pada langkah 2 dimasukkan dalam fungsi aktivasi untuk menghasilkan nilai output (y'). Fungsi aktivasi yang sering digunakan adalah fungsi aktivasi sigmoid dengan persamaan sebagai berikut:

$$y = \frac{1}{1 + \exp(-s)} \quad (2)$$

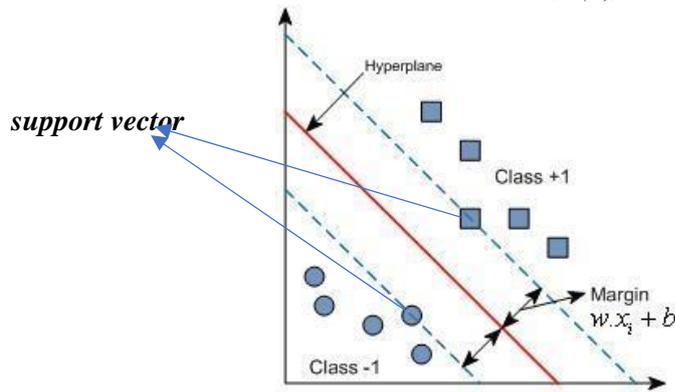
- Menghitung nilai error dengan menggunakan persamaan sebagai berikut:
 $e = y - y'$
- Jika target error yang diinginkan belum tercapai, lakukan perubahan pada bobot.
- Perubahan bobot dihitung dengan mengganti bobot sebelumnya dengan $w(k) + \eta e x_i$. Dimana $w(k)$ adalah bobot saat itu dan η adalah konstanta *learning rate*.

2.2 Support Vector Machines (SVM)

Support Vector Machine (SVM) pertama kali diperkenalkan oleh Vapnik pada tahun 1992 sebagai rangkaian harmonis konsep-konsep unggulan dalam bidang *pattern recognition*. *Support Vector Machines* (SVM) menurut Burges (1998) adalah algoritma supervised learning yang sangat kuat yang digunakan untuk klasifikasi atau regresi. Secara lebih dalam, *Support Vector Machines* (SVM) adalah metode klasifikasi yang bekerja dengan cara mencari *hyperplane* dengan margin maksimum. *Hyperplane* bisa berupa garis linear maupun non-linear tergantung pada kondisi data. *Hyperplane* didapat dari margin, jarak terdekat titik-titik data disetiap kelas dari *hyperplane*. Setiap titik data yang menjadi margin disebut *support vector*. *Support vector* yang mempengaruhi bentuk *hyperplane* (fungsi pemisah), sehingga SVM dapat melakukan klasifikasi data dengan persamaan sebagai berikut:

$$f(x, w, b) = \text{sign}(w^T x + b) \quad (3)$$

Pada metode klasifikasi SVM dilakukan dengan mencari *hyperplane* terbaik yang berfungsi sebagai pemisah dua kelas data. Untuk mencari *hyperplane* terbaik adalah dengan mengukur margin *hyperplane* dan mencari titik maksimumnya. Margin adalah jarak antara *hyperplane* dengan *pattern* terdekat dari masing-masing kelas. *Pattern* yang paling dekat ini disebut sebagai *support vector*.



Gambar 1. Struktur SVM Pembagian Dua Kelas (+1 dan -1) Berdasarkan Garis *Hyperplane*

Berdasarkan persamaan (3) dan Gambar 1, jika $w x_1 + b = +1$ adalah *hyperplane* pendukung dari kelas +1 dan $w x_2 + b = -1$ adalah *hyperplane* pendukung dari kelas -1, maka margin dapat dihitung dengan mencari jarak dua *hyperplane* pendukung. Sehingga

$$(w x_1 + b = +1) - (w x_2 + b = -1) \Rightarrow w(x_1 - x_2) = 2$$

$$\left(\frac{\mathbf{w}}{\|\mathbf{w}\|} (x_1 - x_2) \right) = \frac{2}{\|\mathbf{w}\|} \quad (4)$$

Margin terbesar dapat ditemukan dengan memaksimalkan nilai jarak antara *hyperplane* dan titik terdekatnya, yaitu $1 / \|\mathbf{w}^{\rightarrow}\|$. Hal ini dapat dirumuskan sebagai *Quadratic Programming* (QP) Problem. Untuk klasifikasi linier di dalam *primal space*, formulasi optimasi SVM adalah sebagai berikut:

$$\frac{1}{2} \min (\|\mathbf{w}\|^2) \quad (5)$$

Dengan syarat $y_i (\mathbf{x}^T \mathbf{w} + b) \geq 1, i=1,2, \dots, N$

Permasalahan ini dapat dipecahkan dengan berbagai, salah satunya dengan *Lagrange Multiplier*.

$$L(\mathbf{w}, b, \alpha) = \frac{\|\mathbf{w}\|^2}{2} - \sum_{i=1}^l \alpha_i (y_i (\mathbf{x}^T \mathbf{w} + b) - 1) \quad (i = 1,2,3,\dots,l) \quad (6)$$

α_i adalah *Lagrange Multiplier*, yang bernilai nol atau positif ($\alpha_i \geq 0$). Nilai optimal dari persamaan diatas dapat dihitung dengan meminimalkan L terhadap \mathbf{w}^{\rightarrow} dan b, dan memaksimalkan L terhadap α_i . Untuk menyelesaikan problem *non-linear*, SVM dimodifikasi dengan memasukkan fungsi kernel. Dalam *non-linear SVM*, pertama-tama data \vec{x} dipetakan oleh fungsi $\Phi(\vec{x})$ ke ruang vektor yang berdimensi lebih tinggi. *Hyperplane* yang memisahkan kedua kelas tersebut dapat dikonstruksikan. Selanjutnya gambar (1) menunjukkan bahwa fungsi Φ memetakan tiap data pada input space tersebut

ke ruang vektor baru yang berdimensi lebih tinggi (dimensi 3), sehingga kedua kelas dapat dipisahkan secara linear oleh sebuah *hyperplane*.

Fungsi kernel yang umum digunakan pada metode SVM adalah sebagai berikut:

Linier	$K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}^T \mathbf{x}_j \mathbf{i}$
Polynomial	$K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma \mathbf{x}^T \mathbf{x}_j + r)^p, \gamma > 0 \mathbf{i}$
Radial Basis Function (RBF)	$2 K(\mathbf{x}_i, \mathbf{x}_j) = \exp [-\gamma \ \mathbf{x}_i - \mathbf{x}_j\] , \gamma > 0$
Sigmoid	$K(\mathbf{x}_i, \mathbf{x}_j) = \tanh (\gamma \mathbf{x}^T \mathbf{x}_j + r) \mathbf{i}$

2.3 Evaluasi dan Validasi

Cross Validation merupakan salah satu metode statistik yang digunakan untuk menilai/memvalidasi keakuratan sebuah model yang dibangun berdasarkan dataset tertentu. *Validation* juga merupakan pengujian standar yang dilakukan untuk memprediksi *error rate*. Untuk evaluasi dan validasi model pada penelitian ini menggunakan metode *K-Fold Cross Validation* (K-Fold=10) sehingga akan didapatkan tingkat akurasi, *class precision*, *recall*, dan nilai AUC. Kriteria-kriteria tersebut yang akan dipakai dalam membandingkan model NN dan SVM.

HASIL DAN PEMBAHASAN

Korea Selatan telah berhasil menyebarluaskan budayanya melalui industri hiburan seperti *K-drama* dan musik *K-Pop*, menciptakan *Korean Wave* di seluruh dunia. Budaya Korea Selatan, termasuk *game online*, kosmetik, *fashion*, animasi, dan makanan, semakin dikenal dan dipelajari dalam era *Hallyu 3.0*. Salah satunya Indonesia, dimana minat masyarakat Indonesia terhadap budaya Korea Selatan sangatlah tinggi. Menurut portal berita Tempo, pada rentang tahun 2020 hingga 2021, dari 7,5 miliar cuitan yang membahas mengenai *K-Pop*, Indonesia menjadi negara dengan cuitan mengenai *K-Pop* terbanyak di dunia. Indonesia juga turut menduduki peringkat pertama daftar negara dengan jumlah penggemar *K-Pop* di negara asalnya, yaitu Korea Selatan.

3.1 Prapemrosesan data

Pada dataset yang digunakan terdapat nilai-nilai yang kosong (*missing value*), sehingga pada penelitian ini untuk mengatasi nilai-nilai yang kosong tersebut dilakukan *cleansing* terhadap data yaitu dengan cara mereduksi tuple yang memiliki nilai kosong.

3.2 Modelling

Pada tahap *modelling* akan dilakukan klasifikasi terhadap data. Sebelum dilakukan klasifikasi, terlebih dahulu dilakukan *split* data yaitu membagi data menjadi subset *data training* dan *data testing*. Adapun tujuan dari pembagian subset data tersebut adalah menggunakan *data training* sebagai dasar pembuatan model dan menggunakan *data testing* sebagai dasar evaluasi atas diterapkannya model pada data eksisting.

3.3 Split Data

Proses *split data* dalam penelitian ini menggunakan teknik *K-Fold Cross Validation*. Proses pembagian data *K-Fold Cross Validation* memiliki mekanisme menjadikan $(1 - (1/k))$ bagian data menjadi *data training* dan $(1/k)$ bagian data lainnya menjadi *data testing*. *K-Fold Cross Validation* yang digunakan dalam penelitian ini adalah 10 fold CV dengan pertimbangan kriteria 10 fold CV telah direkomendasikan dalam pemilihan model terbaik karena cenderung memberikan estimasi akurasi yang kurang bisa dibandingkan dengan CV nilai lainnya [10].

3.4 Klasifikasi

Proses klasifikasi dalam penelitian ini menggunakan *software* RapidMiner Studio sebagai alat bantu. Arsitektur proses ANN dalam penelitian ini ditetapkan menggunakan 1 *hidden layer* dan 7 *neuron*. Parameter ANN yang digunakan adalah *learning rate* dan *momentum*. Parameter *learning rate* ditetapkan pada range 0,01 sampai 0,99 dengan *steps* sebanyak 5 dan menggunakan *scale linear*. Sedangkan parameter *momentum* ditetapkan pada range 0,01 sampai 0,9 dengan *steps* dan *scale* yang sama dengan parameter *learning rate*. Dalam proses SVM, fungsi Kernel yang digunakan adalah *Radial Basis Function* (RBF). Pada fungsi Kernel RBF ditetapkan parameter gamma dan C. Nilai gamma dan C yang digunakan dalam penelitian ini adalah 0,1, 1,0 dan 10.

3.5 Evaluasi

Pada tahap evaluasi akan dilakukan evaluasi ukuran *precision* dan *recall* dari model. *Precision* merupakan ukuran ketepatan model klasifikasi dalam mengklasifikasi data yang diinginkan/relevan atau dapat dikatakan merupakan ukuran dari kualitas model klasifikasi. *Recall* merupakan ukuran seberapa lengkap model klasifikasi dapat mendeteksi data yang diinginkan/relevan atau dapat dikatakan merupakan ukuran dari kuantitas model klasifikasi.

Artificial Neural Network (ANN)

Dengan menggunakan dataset yang telah melalui tahap sebelumnya, langkah selanjutnya yaitu mencari model terbaik dengan menggunakan ANN. Nilai parameter Neural Network dilakukan dengan cara optimasi parameter dengan menggunakan operator *optimized parameter (grid)* pada *software* RapidMiner. Selanjutnya, proses klasifikasi menghasilkan ukuran-ukuran *performance* yang dapat dijadikan acuan evaluasi pemilihan model klasifikasi yang baik untuk data.

Tabel 2. Hasil AUC dan Akurasi ANN

	AUC	Akurasi Rate	Learning Rate	Mome ntum
ANN	0,834	77,60	0,01	0,420

Berdasarkan Tabel 2 ANN memiliki nilai AUC sebesar 0,834 dan akurasi sebesar 77,60 %, atau dapat diartikan bahwa ANN mampu mengklasifikasikan dengan benar orang menderita diabetes dan tidak menderita diabetes sebesar 77,60%.

Tabel 3. Ukuran Kinerja Klasifikasi ANN

	true Yes	true No	class precision
pred. Yes	163	67	70,87%
pred. No	105	433	80,48%
class recall	60,82%	86,60%	

Kinerja klasifikasi ANN dapat dilihat dari Tabel 3. Nilai *class recall* “Yes” sebesar 60,82% dapat diartikan bahwa model mampu secara tepat mengklasifikasikan 163 penderita diabetes pada data aktual sedangkan 105 orang *mismatch* dan diklasifikasikan dalam bukan penderita diabetes. Nilai *class recall* “No” sebesar 86,60% dapat diartikan bahwa model mampu secara tepat mengklasifikasikan 67 penderita diabetes pada data aktual sedangkan 433 orang *mismatch* dan diklasifikasikan dalam bukan penderita diabetes. Nilai *class precision* sebesar 70,87%, artinya model dapat memprediksi penderita diabetes dengan tepat sebesar 70,87% dan memprediksi bukan penderita diabetes dengan tepat sebesar 80,48%.

Support Vector Machines (SVM)

Nilai parameter SVM dilakukan dengan cara optimasi parameter dengan menggunakan operator *optimized parameter (grid)* pada *software* RapidMiner. Selanjutnya, proses klasifikasi menghasilkan ukuran-ukuran *performance* yang dapat dijadikan acuan evaluasi pemilihan model klasifikasi yang baik untuk data.

Tabel 4. Hasil AUC dan Akurasi SVM

	AUC	Akurasi	Gamma	C
ANN	0,688	65,24	0,1	10

Berdasarkan Tabel 4 SVM memiliki nilai AUC sebesar 0,688 dan akurasi sebesar 65,24 %, atau dapat diartikan bahwa SVM mampu mengklasifikasikan dengan benar orang menderita diabetes dan tidak menderita diabetes sebesar 65,24%. Selanjutnya adalah melihat ukuran kinerja klasifikasi SVM kernel RBF melalui *confusion matrix* yang ditampilkan pada Tabel 5 berikut.

Tabel 5. Ukuran Kinerja Klasifikasi SVM

	true Yes	true No	class precision
<i>PERBANDINGAN METODE NEURAL NETWORK DAN SUPPORT VECTOR MACHINE DALAM KLASIFIKASI DIAGNOSA PENYAKIT DIABETES</i>			

pred. Yes	1	0	100,00%
pred. No	267	500	65,19%
class recall	0,37%	100,00%	

Tabel 5 menunjukkan bahwa model SVM Kernel RBF memiliki nilai *class recall* “Yes” sebesar 0,37% dapat diartikan bahwa model mampu secara tepat mengklasifikasikan 1 penderita diabetes pada data aktual sedangkan 267 orang *mismatch* dan diklasifikasikan dalam bukan penderita diabetes. Nilai *class recall* “No” sebesar 100% dapat diartikan bahwa model mampu secara tepat mengklasifikasikan 500 orang yang tidak menderita diabetes sesuai data aktual. Nilai *class precision* sebesar 100%, artinya model dapat memprediksi penderita diabetes dengan tepat sebesar 100% dan memprediksi bukan penderita diabetes dengan tepat sebesar 65,19%.

Berdasarkan evaluasi pemilihan model klasifikasi ANN dan SVM Kernel RBF, model ANN mampu mendapatkan nilai AUC sebesar 0,834 dengan tingkat akurasi sebesar 77,60%. Oleh karena itu, berdasarkan nilai AUC model ANN dapat dikategorikan dalam *good classification*. Model SVM memiliki nilai AUC sebesar 0,688 dengan akurasi sebesar 65,24% dan dapat dikategorikan dalam *poor classification*.

KESIMPULAN

Berdasarkan penelitian yang sudah dilakukan guna memprediksi penyakit diabetes menggunakan metode ANN dan SVM, maka kesimpulan yang bisa diambil antara lain adalah sebagai berikut: 1. Berdasarkan hasil analisis pada penelitian ini didapatkan nilai akurasi untuk metode ANN sebesar 77,60%, sedangkan nilai akurasi untuk metode SVM sebesar 65,24%. Artinya penggunaan metode ANN lebih baik daripada SVM untuk mengklasifikasikan seseorang menderita diabetes atau tidak. 2. Pada metode ANN memiliki nilai AUC sebesar 0,834 sehingga dapat dikategorikan dalam *good classification*.

REFERENSI

- ADA. American Diabetes Association Standards Of Medical Care In Diabetes, the journal Of Clinical And Applied Research And Education, USA, Volume 38, Supplement 1, ISSN 0149-5992. 2015.
- Burges, C. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2, 2. 1998.
- Fausett, L. V. *Fundamentals of neural networks*. Prentice-Hall. 1994.
- Haykin. S. *Neural Networks and Learning Machines*. United State of America: Pearson. 2009.
- Kementerian Kesehatan. *Potret Sehat Indonesia dari Riskesdas 2018*, 2 November 2018.
<<https://www.kemkes.go.id/article/view/18110200003/potret-sehat-indonesia-dari-riskesdas-2018.html>> [diakses pada 6 Mei 2021].
- Larose, Daniel T. *Discovering Knowledge In Data: An Introduction to Data mining*. New Jersey: John Willey & Sons. Inc. 2005.
- M.B. Hanif, Khoirudin, “*Sistem Aplikasi Prediksi Penyakit Diabetes Menggunakan Fitur Selection Korelasi Pearson dan Klasifikasi Naïve Bayes*”, *PERBANDINGAN METODE NEURAL NETWORK DAN SUPPORT VECTOR MACHINE DALAM KLASIFIKASI DIAGNOSA PENYAKIT DIABETES*

- Pengembangan Rekayasa dan Teknologi, Vol. 16 No.2 (Hal 199-205), 2020.
- Nisbet, Robert, John Elder, dan Gary Miner. Handbook of Statistical Analysis and Data Mining Applications. California: Elsevier Inc. 2009.
- PERKENI. Konsensus Pengelolaan dan Pencegahan Diabetes Melitus Tipe 2 di Indonesia. Jakarta: PERKENI; 2011.
- P.R. Sihombing, O.P. Hendarsin, “*Perbandingan Metode Artificial Neural Network (ANN) dan Support Vector Machine (SVM) untuk Klasifikasi Kinerja Perusahaan Daerah Air Minum (PDAM) di Indonesia*”, Jurnal Ilmu Komputer Vol. XIII No.1 (Hal 9-20), 2018.
- Smeltzer & Bare . Textbook of Medical Surgical Nursing Vol.2. Philadelphia: Lippincott William & Wilkins. 2008.
- UCI Machine Learning. “Pima Indians Diabetes Database-Predict the onset of diabetes based on diagnostic measures”, 7 Oktober 2016. <<https://www.kaggle.com/uciml/pima-indians-diabetes-database/metadata>> [dikases pada 1 Mei 2021].
- WHO. Global Report On Diabetes. France: World Health Organization; 2016.



© 2021 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY SA) license (<https://creativecommons.org/licenses/by-sa/4.0/>).