INFORMATICS AND DIGITAL EXPERT (INDEX) - VOL. 4 NO. 1 (2022) 40-45
Terbit *online* pada laman web jurnal : https://e-journal.unper.ac.id/index.php/informatics
| ISSN (Print) 2775-2208 | ISSN (Online) 2715-0453 |

Prediksi Penyakit Diabetes Menggunakan Algoritma *Support Vector Machine* (SVM)

*Hovi Sohibul Wafa[1], Asep Id Hadiana[*2], Fajri Rakhmat Umbara[*3]*

[123] *Informatika, Jl. Terusan Jend Sudirman, Cibeber, Kec. Cimahi Sel., Kota Cimahi, 40531, Indonesia*
E-mail: hovi.sohibul@student.unjani.ac.id[1], asep.hadianan@lecturer.unjani.ac.id[*2], fajri.umbara@lecturer.unjani.ac.id[*3]

INFORMASI ARTIKEL

Sejarah Artikel:
Diterima Redaksi : 18 Februari 2022
Revisi Akhir : 13 Juni 2022
Diterbitkan *Online* : 03 Juli 2022

Kata Kunci:
Diabetes Mellitus, Prediksi, Support Vector Machine, K-Fold, Confusion Matrix, Mean Square Error

Korespondensi:
Telepon / Hp : +62 (085) 797011634
E-mail :
hovi.sohibul@student.unjani.ac.id

A B S T R A K

Diabetes Mellitus (DM) atau lebih dikenal dengan sebutan penyakit kencing manis adalah penyakit kronis yang disebabkan oleh gagalnya organ pankreas memproduksi jumlah hormon insulin secara memadai sehingga menyebabkan peningkatan kadar glukosa dalam darah. Diabetes Mellitus merupakan penyakit yang berbahaya, banyak diberbagai negara terkena penyakit diabetes termasuk di Indonesia. Penyebab utama diabetes masih belum diketahui, namun banyak yang percaya bahwa faktor genetika dan gaya hidup dapat memainkan peran utama pada diabetes. Para peneliti di bidang bioinformatika telah berusaha untuk mengatasi penyakit ini dan membuat sistem untuk membantu dalam prediksi diabetes. Dari berbagai penelitian yang ada, banyak menggunakan metode seperti C4.5, KNN, Naïve Bayes, serta SVM Linier dalam membangun sistem, tapi metode SVM *Radial Basis Function* (RBF) jarang digunakan dikarenakan hasil akurasi yang didapat tidak cukup untuk digunakan pada sistem prediksi diabetes. Pada penelitian ini menjawab gap tersebut bahwa dengan menggunakan metode algoritma SVM *Radial Basis Function* (RBF) dapat menghasilkan akurasi yang tinggi dengan mencapai sebesar 91%. Pengujian akurasi yang dilakukan menggunakan Confusion Matrix dan peramalan Mean Square Error dengan kfold kelipatan 10. Penelitian ini bertujuan untuk menentukan apakah penderita/pasien dapat terkena penyakit diabetes atau tidak dengan menerapkan teknik data mining dan klasifikasi menggunakan algoritma SVM *Radial Basis Function* berbasis *Forward Selection*.

1. PENDAHULUAN

Diabetes Melitus tipe dua adalah suatu penyakit atau gangguan metabolisme kronis dengan multi etilogi yang ditandai dengan tingginya kadar gula disertai dengan gangguan metabolisme karbohidrat, lipid, dan protein sebagai akibat insufisien fungsi insulin. Insufisiensi fungsi insulin dapat disebabkan oleh gangguan atau defisiensi produksi insulin oleh sel-sel beta *langerhans* kelenjar pankreas, atau disebabkan oleh kurangnya responsifnya sel-sel tubuh terhadap insulin [1]. Insulin merupakan hormon yang dihasilkan pankreas untuk membantu mengendalikan gula darah dan membantu penyerapan glukosa ke dalam sel-sel tubuh untuk mengendalikan gula darah. Glukosa sendiri biasanya berasal dari makanan yang mengandung karbohidrat, dan diubah tubuh menjadi sumber energi. Hormon ini berkaitan erat dengan masalah kesehatan yang menyebabkan kadar gula darah tinggi (*hyperglycemia*) dan gula darah rendah (*hypoglecemia*) [2]. Penyakit Diabetes Melitus tipe 2 dikenal sebagai penyakit *silent killer* dikarenakan para penderita tidak menyadari dan saat diketahui sudah terjadi komplikasi. Dari beberapa penelitian sebelumnya yang meneliti mengenai penyakit diabetes tipe 2 [3] [4], menjelaskan bahwa hasil dari penelitian tersebut mengungkapkan kebanyakan masyarakat terkena diabetes tipe 2. Hal tersebut dikarenakan pola hidup yang tidak teratur dan konsumsi makanan yang mengandung gula secara berlebihan.

Menurut Organisasi Internasional Diabetes Federation (IDF), memperkirakan sedikitnya terdapat 483 juta orang pada usia 20-79 tahun di dunia menderita diabetes pada tahun 2019 atau setara dengan angka prevalensi sebesar 9.3% dari total penduduk pada usia yang sama. Berdasarkan jenis kelamin, IDF memperkirakan prevalensi diabetes pada tahun 2019 yaitu 9% pada perempuan dan 9.65% pada laki-laki. Prevalensi diabetes diperkirakan meningkat seiring penambahan umur penduduk menjadi 19.9% atau 112 juta orang pada umur 65-79 tahun. Angka ini diprediksi terus meningkat hingga mencapai 578 juta di tahun 2030 dan 700 juta di tahun 2045 [5].

Penyebab utama dari penyakit ini masih belum diketahui, namun para dokter berasumsi bahwa pada kondisi tersebut diduga berkaitan dengan gaya hidup dan pola makan memainkan peran utama dalam penyakit tersebut. Penderita diabetes menghadapi resiko terkena beberapa masalah kesehatan sekunder seperti penyakit jantung, kerusakan saraf, dll [6]. Untuk mengendalikan peningkatan terkena diabetes tipe 2, perlunya mendiagnosis penyakit tersebut secara dini untuk mencegah komplikasi dan mengurangi resiko masalah kesehatan yang terbilang parah [7] [8]. Namun, para dokter perlu menganalisa beberapa faktor sebelum mendiagnosis yang membuat pekerjaan dokter menjadi sulit. Akan tetapi, dari perkembangan zaman sekarang hal tersebut dapat dilakukan dengan adanya teknologi untuk membuat suatu prediksi/mendeteksi penyakit diabetes [9]. Dari perkembangan teknologi tersebut

dapat membantu meringankan pekerjaan dokter dalam memprediksi penyakit tersebut.

Yang diperlukan dalam membangun teknologi untuk mengolah data diabetes tipe 2 menjadi sebuah prediksi yaitu dengan teknik *data mining*. *Data mining* sebagai proses seleksi, eksplorasi, dan pemodelan dari sejumlah data besar untuk menghasilkan sebuah pengetahuan [10]. Salah satu teknik *data mining* yang digunakan dalam suatu prediksi adalah klasifikasi. Tugas klasifikasi adalah memprediksi dengan keluaran *variable/class* yang bernilai kategorial atau polinomial [11]. Alhasil dengan teknik tersebut menghasilkan sebuah sistem yang dapat memprediksi penyakit tersebut. Dari penelitian terdahulu banyak yang menggunakan solusi ini dengan berbagai metode algoritma. Dari menggunakan deep learning yaitu Neural Network [12] hingga menggunakan machine learning seperti C4.5, *Naive Bayes*, *Logistic Regression* dan SVM *Linear* [13] [14] [15]. Akan tetapi dari semua metode algoritma yang digunakan, belum ada penelitian dengan algoritma SVM *Radial Basis Function* dan metode *Forward Selection* dalam memprediksi diabetes. Oleh sebab itu, penelitian yang dilakukan akan menjawab gap tersebut bahwa dengan menggunakan metode algoritma SVM *Radial Basis Function* (RBF) dengan bantuan *Forward Selection* dapat menghasilkan akurasi yang tinggi.

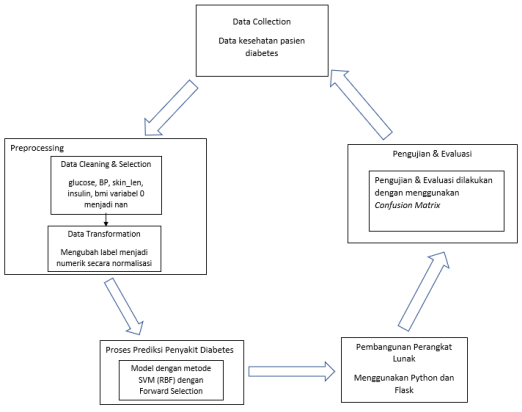
Perlu diketahui bahwa algoritma SVM (*Support Vector Machine*) adalah jenis algoritma lain dari teknik *machine learning*. Algoritma tersebut dikenal salah satu metode klasifikasi yang memiliki hasil tinggi dalam melakukan prediksi pengklasifikasian potensi pada data. Pada algoritma SVM memiliki 2 metode yaitu regresi (*Support Vector Regression*) dan klasifikasi (*Support Vector Classification*). Namun SVM memiliki kelemahan yaitu tidak dapat menemukan pemisah dalam *hyperplane* sehingga tidak memiliki akurasi yang besar dan tidak dapat menggeneralisasi dengan baik [16]. Oleh karena itu, dibutuhkan kernel untuk meningkatkan akurasi ke ruang dimensi yang lebih tinggi yang disebut ruang kernel. *Radial Basis Function* (RBF) adalah salah satu kernel yang dianggap memiliki akurasi tinggi. *Radial Basis Function* (RBF) sebagai neuron dan menggunakannya sebagai cara untuk membandingkan data input dengan data pelatihan. Vektor input diproses oleh beberapa neuron fungsi radial basis, dengan bobot yang bervariasi, dan jumlah total neuron menghasilkan nilai kesamaan. Jika vektor input cocok dengan data pelatihan, akan memiliki nilai kesamaan yang tinggi. Atau, jika tidak cocok dengan data pelatihan, maka tidak akan diberi nilai kesamaan yang tinggi [16] [17].

Metode *Forward Selection* dilakukan dengan cara memasukan prediktor secara bertahap, prediktor ini berdasarkan korelasi parsial terbesar. Dalam metode ini, variabel prediktor yang dimasukan dalam model tidak akan dapat dikeluarkan lagi. Proses tersebut dihentikan ketika prediktor-prediktor baru tidak bisa meningkatkan berpengaruh secara signifikan (sig dibawah 0.05) terhadap variabel respon. Karena itulah metode *forward selection* menjadi salah satu prosedur pemilihan model terbaik dengan eliminasi variabel bebas yang membangun model secara bertahap [18]. Dari penjelasan singkat diatas, penelitian yang dilakukan terkait prediksi penyakit diabetes tipe 2 adalah untuk bisa membantu menekan jumlah penderita diabetes serta mengetahui penyakit tersebut sebelum berkomplikasi secara menyeluruh dengan bantuan metode SVM *Radial Basis Function* (RBF) dan *Forward Selection*. Serta hasil dari algoritma *Support Vector Machine* dievaluasi dengan *k-fold cross validation* dengan nilai cv bernilai 10 serta dilakukan pengujian dengan teknik perhitungan *Confusion Matrix* [19] [20].

Penelitian ini bertujuan untuk menentukan apakah penderita/pasien dapat terkena penyakit diabetes atau tidak dengan menerapkan teknik data mining dan klasifikasi menggunakan algoritma SVM *Radial Basis Function* berbasis *Forward Selection* dan menghitung performa *accuracy*, *precision*, *recall*, dan *f1-score* yang dihasilkan dari pengujian *Confusion Matrix*

2. METODE PENELITIAN

Dalam metode penelitian ini, memiliki tahapan-tahapan yang dilakukan dalam membangun sistem prediksi diabetes ini. Dari mulai pengumpulan data yang digunakan, kemudian melakukan *processing* pada data, selanjutnya proses menambang data mengguanakan metode *support vector machine* dan *forward selection*, lalu implementasi menggunakan bahasa pemograman python, serta evaluasi pengujian dengan *confusion matrix*. Berikut adalah gambarannya.



Gambar 1 Metode Penelitian

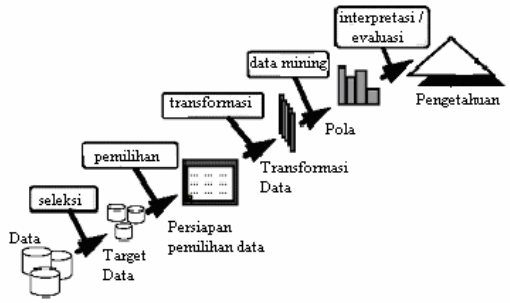
2.1. Data Collection

Data yang digunakan dalam penelitian ini adalah data dengan jenis wanita dengan keturunan indian pima yang memiliki 8 atribut dan 1 label, 8 atribut terdiri dari *n_pregnant*, *glucose_conc*, *bp*, *skin_len*, *insulin*, *bmi*, *pedigree_funct*, *age*, dan label *output*.

2.2. Data Mining

Perlu diketahui bahwa data mining merupakan proses pengumpulan sebuah informasi penting pada suatu data yang berukuran besar. Data mining dapat diartikan sebagai sebuah teknik penambangan data untuk melakukan suatu analisa dengan teknik penyaringan informasi secara lebih akurat. Teknik tersebut biasanya dilakukan untuk menemukan beberapa pola-pola tertentu yang masih memiliki relevansi dengan *goals* [10] [11]. Ada beberapa teknik yang digunakan pada data mining. Diantaranya yaitu teknik klasifikasi. Klasifikasi adalah sebuah proses untuk menemukan model atau fungsi dengan mengelompokan kelas data dengan tujuan untuk memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Teknik ini banyak digunakan dalam berbagai bidang, seperti dibidang kesehatan yaitu prediksi penyakit kanker [21], mendiagnosis riwayat kesehatan pasien [22] [23].

Beberapa tahapan-tahapan yang perlu dilakukan pada data mining. Hal yang pertama adalah seleksi data, pemilihan data sebelum tahap penggalian informasi dimulai. Hasil yang diseleksi digunakan kedalam proses data mining. Kedua *pre-processing*, dalam tahapan ini memiliki 3 step diantaranya data *cleaning*, menghilangkan data-data yang tidak lengkap. Lalu data *selection*, pengambilan data yang relevan/sesuai dengan tugas analisis. Dan data *intergation*, proses penggabungan beberapa data sumber data. Ketiga transformasi data, data yang telah terpilih, akan ditransformasikan ke dalam bentuk yang cocok untuk prosedur penggalian lebih lanjut dengan cara melakukan proses normalisasi dan agregasi. Keempat yaitu menambang data, adalah suatu proses mencari pola atau informasi pada data yang terpilih menggunakan metode tertentu. Terakhir evaluasi, dapat diartikan menampilkan pola informasi yang telah diolah pada tahap data mining kedalam bentuk yang mudah dimengerti [24]. Tahapan tersebut dapat dilihat pada Gambar 1. Tahapan-Tahapan Proses Dalam Data Mining



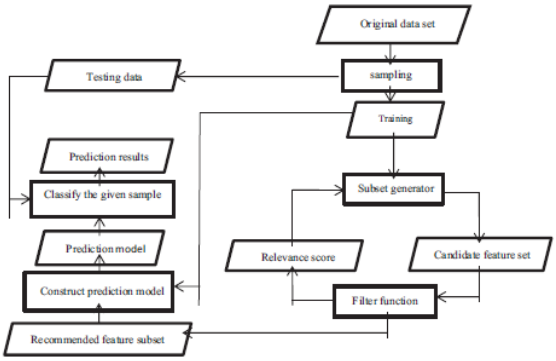
Gambar 2 Tahapan-Tahapan Proses Dalam Data Mining

2.3. Forward Selection

Forward Selection adalah suatu metode dengan bertujuan untuk menambah variabel yang ditentukan satu per satu ke dalam persamaan yang didasarkan pada nilai tertentu untuk masukan. Dimana nilai tersebut dimulai dengan tidak adanya prediktor variabel atau model yang berisi nilai konsta. Lalu memilih variabel dengan nilai p-value dari masing-masing prediktor, jika nilai p tersebut dari prediktor kurang dari nilai taraf nyata α, maka prediktor tersebut tetap dipakai dalam persamaan. Kemudian melakukan pengulangan dengan cara yang sama hingga variabel terakhir tidak memiliki koefisien regresi dan tidak signifikasn [20] [18]. Koefisien regresi yang signifikan dari variabel terakhir dilihat dari uji persaman terakhir.

Gambar 3 Langkah-langkah Klasifikasi dengan Wrapper Method

Dari langkah-langkah yang ada pada gambar diatas,



dirumuskan pada prosedur berikut:

- Menentukan model awal $\hat{y} = b0$
- Memasukkan variable respon dengan setiap variable berprediktor, misalnya $X1, X2, \ldots . Xn$ yang terkait dengan $\hat{y}$. Misalkan $X1$ sehingga membentuk model $\hat{y} = b0 + b1X1$
- Uji F terhadap peubah pertama yang terpilih. Jika $Fhitung < Ftabel$ maka peubah terpilih dibuang dan proses dihentikan. Apabila $Fhitung > Ftabel$ maka peubah terpilih memiliki pengaruh nyata terhadap peubah terkait y, sehingga layak untuk diperhitungkan di dalam model.
- Masukan peubah bebas terpilih (yang paling signifikan) ke dalam model. Misalkan $X2$, sehingga membentuk suatu model $\hat{y} = b0 + b1X1 + b2X2$
- Uji F, jika $Fhitung < Ftabel$ maka proses dihentikan dan model terbaik adalah model sebelumnya. Namun jika $Fhitung \geq Ftabel$, variable peubah bebas layak untuk dimasukan ke dalam model dan kembali ke langkah C. proses akan berakhir jika tidak ada lagi peubah yang tersisa yang bias dimasukkan ke dalam model.

2.4. Support Vector Machine

SVM adalah satu set metode pembelajaran yang diawasi terkait yang digunakan dalam diagnosis medis untuk klasifikasi dan regresi. SVM secara bersamaan meminimalkan kesalahan klasifikasi empiris dan memaksimalkan margin geometrik. Jadi SVM disebut Pengklasifikasi Margin Maksimum.SVM adalah algoritma umum yang didasarkan pada batasan risiko terjamin dari teori pembelajaran statistik yaitu yang disebut prinsip minimisasi risiko struktural. SVM menemukan sebuah hyperplane yang memiliki kemungkinan fraksi poin terbesar dari kelas yang sama pada bidang yang sama. *Hyperplane* adalah fungsi yang digunakan untuk membedakan antar fitur. Dalam 2-D,

fungsi yang digunakan untuk mengklasifikasikan antar fitur adalah garis sedangkan fungsi yang digunakan untuk mengklasifikasikan fitur dalam 3-D disebut sebagai bidang, begitu pula fungsi yang mengklasifikasikan titik dalam dimensi yang lebih tinggi disebut sebagai *hyperplane*. Namun hasil akurasi yang didapatkan masih belum cukup dikategorikan baik dan tidak digeneralisasikan dengan baik. Oleh karena itu perlu meningkatkan akurasi ke dimensi yang lebih tinggi, hal ini disebut sebagai fungsi kernel. Fungsi kernel adalah suatu fungsi k yang mana untuk semua vektor input x, z akan memenuhi kondisi $k(x,z) = \emptyset(x)^t\emptyset(z)$ dimana $\emptyset(.)$ adalah fungsi pemetaan dari ruang input ke ruang fitur. Dengan kata lain, fungsi kernel adalah fungsi perkalian dalam (*inner product*) pada ruang fitur. Fungsi kernel memungkinkan untuk mengimplementasikan suatu model pada ruang dimensi lebih tinggi (ruang fitur) tanpa harus mendefinisikan fungsi pemetaan dari ruang input ke ruang fitur. Sehingga, *hyperplane* dapat digunakan sebagai *decision boundary* secara efisien. Salah satu tipe pada kernel adalah *Radial Basis Function*(RBF) /kernel *Gaussian* [16] [17]. Kernel RBF adalah fungsi yang nilainya tergantung pada jarak dari titik asal atau dari beberapa titik [16]. Sebagai berikut:

$$\boldsymbol{K(x_i, x_j) = exponent\left(-\frac{\|xi-xj\|^2}{2\sigma^2}\right)} \quad \textbf{(1)}$$

2.5. Confusion Matrix

Confusion Matrix adalah metode pengukuran untuk mencari masalah klasifikasi yang dilakukan oleh machine learning dengan keluaran berupa dua kelas atau lebih. *Confusion Matrix* adalah tabel dengan 4 kombinasi berbeda dari nilai prediksi dan nilai aktual. Ada empat istilah yang merupakan representasi hasil proses klasifikasi pada confusion matrix yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FN), dan *False Negative* (FN). Nilai *True Negative* (TN) dikategorikan sebagai data yang bersifat negatif terdeteksi dengan benar, lalu *False Positive* (FP) dikategorikan sebagai data bersifat negatif namun terdeteksi menjadi data bersifat positif. Kemudian, *True Positive* (TP) dikategorikan sebagai data bersifat positif yang terdeteksi benar. *False Negative* (FN) merupakan kebalikan dari *True Positive*, data yang bersifat positif, namun terdeteksi sebagai data negatif [19].

		True Values	
		True	False
Prediction	True	TP Correct result	FP Unexpected result
	False	FN Missing result	TN Correct absence of result

Gambar 4 Confusion Matrix

Dari hasil proses klasifikasi pada *confusion matrix*, dihitung tiap nilai dengan *accuracy*, *precision*, *recall*, dan F-1 *score*.

- Accuracy, menggambarkan seberapa akurat model yang digunakan dalam klasifikasi.

  Rumus:

  $$\boldsymbol{accuracy = \left(\frac{(TP+TN)}{(TP+FP+TN+FN)}\right) \times 100\%} \quad \textbf{(2)}$$

- Precision, menggambarkan akurasi antara data yang diminta dengan hasil prediksi yang diberikan oleh model.

  Rumus:

  $$\boldsymbol{precision = \left(\frac{TP}{(TP+FP)}\right) \times 100\%} \quad \textbf{(3)}$$

- Recall, menggambarkan keberhasilan model dalam menemukan kembali informasi.

  Rumus:

  $$\boldsymbol{recall = \left(\frac{TP}{(TP+FN)}\right) \times 100\%} \quad \textbf{(4)}$$

- F-1 score, menggambarkan perbandingan rata-rata precision dan recall yang dibobotkan

  Rumus:

  $$\boldsymbol{f1 - score = \left(\frac{2 \times precision \times recall}{(precision+recall)}\right) \times 100\%} \quad \textbf{(5)}$$

3. HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini, data yang diambil dari website dengan jumlah record 1017 data. Dimana data tersebut adalah data penderita diabetes tipe 2 keturunan indian pima dengan 8 atribut dan 1 label yaitu *n_pregnant*, *glucose_conc*, *bp*, *skin_len*, *insulin*, *bmi*, *pedigree_funct*, *age*, dan label *output*. Berikut adalah Table Dataset Pima Indian.

Tabel 1 Deskripsi Atribut Data

No	Atribut	Deskripsi
1	n_pregnant	Merupakan jumlah banyaknya melahirkan.
2	glucose_conc	Merupakan konsentrasi glukosa plasma 2 jam dalam tes toleransi glukosa oral.
3	bp	Merupakan tekanan darah diastol(mmHg).
4	skin_len	Merupakan ketebalan pada lipatan kulit trisep (mm).
5	insulin	Merupakan serum insulin selama 2 jam.
6	bmi	Merupakan indeks massa badan (weight in kg/(height in m)^2).
7	pedigree_funct	Merupakan riwayat diabetes melitus pada kerabat keturunan yang memiliki hubungan genetik dengan subjek.
8	age	Merupakan umur pasien.
9	output	Merupakan label yang menandakan YES untuk mengidap diabetes dan NO untuk tidak mengidap diabetes.

Sebelum diolah menggunakan metode SVM dan *Forward Selection*, data terlebih dahulu masuk ke tahap *processing* untuk memastikan bahwa tiap atribut pada data tidak memiliki variabel yang menggangu proses penambangan data melalui proses *cleaning* dan *selection*. Pada penelitian ini, proses *cleaning*, memproses data dengan memilih atribut data dan

membuang apabila tiap atribut pada data memiliki variabel *missing value*(bernilai kosong/tidak lengkap) dan *noise*(tidak tepat). Atribut yang dipilih meliputi *N_pregnant*, *Glucose_conc*, *BP*, *Skin_len*, *Insulin*, *BMI*, *Pedgree_fun*, *Age*, dimana atribut tersebut masih memiliki variabel yang memiliki nilai yang bersifat *noise* terutama pada atribut *glucose_conc*, *bp*, *skin_len*, *insulin*, *bmi*. Pada atribut tersebut memiliki variabel tidak tepat tebanyak. Setelah melalui tahap *cleaning* dari tiap atribut, kemudian melakukan *selection* pada data. *Selection* data yang dimaksud adalah melakukan pemilihan data umtuk dilakukan pengujian dengan membagi data tersebut. Dimana 80% digunakan untuk data latih dan 20% digunakan untuk data uji. Setelah melalui tahap *selection*, data yang akan digunakan di *transformation* ke dalam bentuk yang cocok untuk prosedur penggalian atau disebut *mining*. Proses *tranformation* pada penelitian ini dilakukan dengan proses normalisasi hanya pada label *ouput*. Karena, variabel yang terdapat pada label *output* merupakan data variabel nominal yaitu YES dan NO sedangkan dari atribut lain merupakan data numerik. Oleh sebab itu, pada label *output* dilakukan *transformation* kedalam bentuk data numerik yaitu 1 yang menandakan terkena diabetes tipe 2 dan 0 yang menandakan tidaak terkena diabetes tipe 2. Berikut Tabel 2.

Tabel 2 Hasil Processing

N_pregnant	Glucose_conc	BP	Skin_Len	Insulin	BMI	Pedigree_Fun	Age	Output
6	148.0	72.0	35.0	168.5	33.6	0.627	50	1
1	85.0	66.0	29.0	95.5	26.6	0.351	31	0
8	183.0	64.0	32.0	168.5	23.3	0.672	32	1
1	89.0	66.0	27.0	94.0	28.1	0.167	21	0
0	137.0	75.5	35.0	168.0	43.1	0.467	33	1
5	116.0	74.0	27.0	95.5	25.6	0.201	30	0
3	78.0	50.0	32.0	88.0	31.0	0.248	26	1
10	115.0	70.0	27.0	95.5	35.3	0.134	29	0
2	197.0	70.0	32.0	543.0	30.5	0.158	53	1
8	125.0	96.0	32.0	168.5	34.2	0.232	54	1

Data yang telah melalui tahap *processing*, data kemudian diuji dengan metode *support vector machine* dengan kernel *radial basis function* dan *forward selection*. Data yang digunakan adalah data uji dengan jumlah record 204 record dengan atribut yang sama seperti Tabel 2. Selanjutnya metode *support vector machine* dan *forward selection* akan melakukan klasifikasi terhadap data uji, sehingga dapat memperoleh nilai performansi dari nilai *accuracy*, *precision*, *recall*, dan *f1-score* dengan metode *confusion matrix* dengan nilai 10 untuk *cross validation*. Hasil klasifikasi yang diperoleh *confusion matrix* untuk mencari nilai *True Positive*, *False Positive*, *False Negative*, dan *True Negative*. Dapat dilihat pada Tabel 3.

Tabel 3 Tabel Confusion Matrix

N=204	True	False
True	133	10
False	8	53

Dari hasil klasifikasi, dihitung tiap nilai kedalam rumus *accuracy*, *precision*, *recall*, dan *f1-score*. Berikut hasil performansi pada Tabel 4.

Tabel 4 Hasil Perfomansi Confusion Matrix

Nilai	SVM(RBF)+FS
Accuracy	91.2%
Precision	93.0%
Recall	94.3%
F1-Score	93.7%

Dari hasil yang didapatkan pada tabel diatas. Performa yang didapatkan dari metode *support vector machine* dengan kernel *radial basis function* dan *forward selection* memperoleh hasil sebesar 91.2% untuk *accuracy*, 93.0% untuk *precision*, 94.3% untuk *recall*, dan 93.7% untuk *f1-score*.

1. KESIMPULAN

Berdasarkan pada penelitian ini dibuat menggunakan algoritma *support vector machine* dan *forward selection*. Dimana, 80% dari data digunakan sebagai data latih dan 20% dari data digunakan sebagai data uji untuk memprediksi penyakit diabetes tipe 2 berdasarkan atribut yang ada pada data. Kesimpulannya adalah hasil dari penelitian ini yang dibandingkan dengan penelitian terdahulu dengan metode yang berbeda seperti *neural network* [7] [12], C4.5, *Naive Bayes*, *Logistic Regression* dan SVM *Linier* [13] [14] [15] bahwa dengan menambahkan metode *forward selection*, model *support vector machine* (rbf) mampu memberikan hasil akurasi yang diperoleh sebesar 91.2% untuk *accuracy*, 93.0% untuk *precision*, 94.3% untuk *recall*, dan 93.7% untuk *f1-scorer*, dari hasil evaluasi *confusion matrix*. Hal tersebut menandakan bahwa metode yang digunakan dalam penelitian ini memberikan perfomansi yang cukup baik dalam memprediksi data diabetes.

Saran penelitian selanjutnya adalah penelitian ini mampu mendapat akurasi yang lebih akurat dari data diabetes tipe 2 apabila ada menambahkan data latih ataupun data uji dan saat variabel dari data tersebut berubah

DAFTAR PUSTAKA

[1] D. W. Hestiana, "FAKTOR-FAKTOR YANG BERHUBUNGAN DENGAN KEPATUHAN DALAM PENGELOLAAN DIET PADA PASIEN RAWAT JALAN DIABETES MELLITUS DI KOTA SEMARANG," *Jurnal of*

Health Education, vol. 3, no. 12, pp. 138-145, 2017.

[2] U. Hasanah, "INSULIN SEBAGAI PENGATUR KADAR GULA DARAH," *Jurnal Keluarga Sehat Sejahtera,* vol. 11, no. 22, pp. 42-49, 2013.

[3] N. A. W. S. F. Peter Piko, "Impact of Genetic Factors on the Age of Onset for Type 2 Diabetes Mellitus in Addition to the Conventional Risk Factors," *MDPI,* vol. 11, no. 6, pp. 1-17, 2021.

[4] E. M. G. R. G. D. P. Anna Izzo, "A Narrative Review on Sarcopenia in Type 2 Diabetes Mellitus: Prevalence and Associated Factors," *MDPI,* vol. 13, no. 183, pp. 1-18, 2021.

[5] S. Pangribowo, "Tetap Produktif, Cegah, dan Atasi Diabetes Melitus," *InfoDatin,* 18 Febuari 2020.

[6] L. A. R. A. T. S. Souad Larabi-Marie-Sainte, "Current Techniques for Diabetes Prediction: Review and Case Study," *MDPI: Journal Applied Science,* vol. 9, no. 2, pp. 1-18, 2019.

[7] S. R. S. N. T. H. N. A. M. Jajang Jaya Purnama, "Analisis Algoritma Klasifikasi Neural Network Untuk Diagnosis Penyakit Diabetes," *IJCIT (Indonesian Journal on Computer and Information Technology),* vol. 5, no. 1, pp. 1-7, 2019.

[8] S. A. T. H. K. S. A. A. Dewi Rahma Ente, "KLASIFIKASI FAKTOR-FAKTOR PENYEBAB PENYAKIT DIABETES MELITUS DI RUMAH SAKIT UNHAS MENGGUNAKAN ALGORITMA C4.5," *Indonesian Journal of Statistics and Its Applications,* vol. 4, no. 1, pp. 80-88, 2020.

[9] D. V. V. Aishwarya Mujumdar, "Diabetes Prediction using Machine Learning Algorithms," *ELSEVIER,* vol. 16, no. 5, pp. 292-299, 2019.

[10] T. S. Aline Embun Pramadhani, "PENERAPAN DATA MINING UNTUK KLASIFIKASI PREDIKSI PENYAKIT ISPA (Infeksi Saluran Pernapasan Akut) DENGAN ALGORITMA DECISION TREE (ID3)," *Jurnal Sarjana Teknik Informatika,* vol. 2, no. 1, pp. 831-839, 2014.

[11] A. Saifudin, "METODE DATA MINING UNTUK SELEKSI CALON MAHASISWA PADA PENERIMAAN MAHASISWA BARU DI UNIVERSITAS PAMULANG," *ResearchGate,* vol. 10, no. 1, pp. 25-36, 2018.

[12] M. M. I. Safial Islam Ayon, "Diabetes Prediction: A Deep Learning Approach," *Modern Education and Computer Science Press(MECS),* vol. 2, no. 1, pp. 21-27, 2019.

[13] D. N. J.-B. C. F. Y. W. A. S. R. Y. L. A. G. M. A. D. R. U. V. I. Y. D. W. S. G. H. M. Cheng-Hong Yang, "Prediction of Mortalityinthe Hemodialysis Patient with Diabetes using Support Vector Machine," *revistaclinicapsicologica,* vol. XXIX, no. 4, pp. 219-232, 2020.

[14] P. S. M. P. L. S. P. Arianna Dagliati, "Machine Learning Methods to Predict Diabetes Complications," *SAGE,* vol. 12, no. 2, pp. 295-302, 2018.

[15] D. S. S. Deepti Sisodia, "Prediction of Diabetes using Classification Algorithms," *ELSEVIER,* vol. 13, no. 2, pp. 1578-1585, 2018.

[16] A. Kowalczyk, SUPPORT VECTOR MACHINES, Morrisville: Syncfusion, 2017.

[17] B. M. E. Matthias Ring, "An approximation of the Gaussian RBF kernel for efficient classification with SVM," *ELSEVIER,* vol. 84, no. 3, pp. 107-113, 2016.

[18] A. N. S. Laily Hermawanti, "PENGGABUNGAN ALGORITMA FORWARD SELECTION DAN K-NEAREST NEIGHBOR UNTUK MENDIAGNOSIS PENYAKIT DIABETES DI KOTA SEMARANG," *Momentum,* vol. 12, no. 2, pp. 28-31, 2016.

[19] Q. L. Y. D. ,. S. M. Xinyang Deng, "An improved method to construct basic probability assignment based on the confusion matrix for classification problem," *ELSEVIER,* Vols. 340-341, no. 16, pp. 250-261, 2016.

[20] D. F. Ghebyla Najla Ayuni, "Penerapan Metode Regresi Linear Untuk Prediksi Penjualan Properti pada PT XYZ," *Jurnal Telematika,* vol. 14, no. 2, pp. 79-86, 2019.

[21] H. Harafani, "Forward Selection pada Support Vector Machine untuk Memprediksi Kanker Payudara," *Jurnal Infortech,* vol. 1, no. 2, pp. 131-139, 2019.

[22] A. K. S. S. A. G. H. A.-R. A. T. C. Ali Kalantari, "Computational intelligence approaches for classification of medical data: State-of-the-art, future challenges and research directions," *ELSEVIER,* vol. 276, no. 1, pp. 2-22, 2018.

[23] A. R. Febie Elfaladonna, "ANALISA METODE CLASSIFICATION-DECISSION TREE DAN ALGORITMA C.45 UNTUK MEMPREDIKSI PENYAKIT DIABETES DENGAN MENGGUNAKAN APLIKASI RAPID MINER," *SINTECH,* vol. 2, no. 1, pp. 10-17, 2019.

[24] R. R. Rerung, "Penerapan Data Mining dengan Memanfaatkan Metode Association Rule untuk Promosi Produk," *JTERA - Jurnal Teknologi Rekayasa,* vol. 3, no. 1, pp. 89-98, 2018.

[25] A. R. Febie Elfaladonna, "ANALISA METODE CLASSIFICATION-DECISSION TREE DAN ALGORITMA C.45 UNTUK MEMPREDIKSI PENYAKIT DIABETES DENGAN MENGGUNAKAN APLIKASI RAPID MINER," *SINTECH JOURNAL,* vol. 2, no. 1, pp. 10-17, 2019.