NASKAH PUBLIKASI

PENELITIAN

IMPLEMENTASI *SUPPORT VECTOR MACHINE* (SVM) UNTUK KLASIFIKASI PEDERITA DIABETES MELLITUS

Oleh:

Favorisen R. Lumbanraja, M. Si., Ph. D
Fanni Lufiana, S. Kom.
Yunda Heningtyas, M. Kom.
Dr. Ir. Kurnia Muludi, M. S.Sc.

Diterbitkan pada:

JURNAL KOMPUTASI

E-ISSN 2541-0350

JURUSAN ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
2022

HALAMAN PENGESAHAN

Nama	: Yunda Heningtyas, M.Kom.
NIP	: 19890108 201903 2 014
Jurusan	: Ilmu Komputer
Fakultas	: Matematika dan Ilmu Pengetahuan Alam
Judul Makalah	: Implementasi *Support Vector Machine* (SVM) Untuk Klasifikasi Pederita Diabetes Mellitus
Oleh	: Favorisen R. Lumbanraja, Fanni Lufiana, Yunda Heningtyas, Kurnia Muludi
Dimuat pada	: Jurnal Komputasi Vol. 10 No. 1 April 2022; E-ISSN 2541-0350 Halaman 75-83
Diterbitkan oleh	: Jurnal Komputasi ©2022 Ilmu Komputer Unila Publishing
Diunggah pada	: https://jurnal.fmipa.unila.ac.id/komputasi/article/view/2940

Bandar Lampung, 28 Oktober 2022
Pengusul,

Yunda Heningtyas, M.Kom.
NIP. 19890108 201903 2 014

Mengetahui,
Ketua Jurusan Ilmu Komputer
FMIPA Unila

Didik Kurniawan, S.Si, M.T.
NIP. 19800419 200501 1 004

Dekan FMIPA Unila

Dr. Eng. Suripto Dwi Yuwono, M.T.
NIP. 19740705 200003 1 001

Menyetujui,
Ketua LPPM Universitas Lampung

Prof. Dr. Ir. Lusmeilia Afriani, D.E.A.
NIP.19650510 199303 2 008

ISSN 2541-0350

Volume 10 No.1 2022

JURNAL KOMPUTASI

Ilmu Komputer
Universitas Lampung
Bandar Lampung, April 2022

Jurnal Komputasi
P-ISSN 2541-0296
E-ISSN 2541-0350

VOL 10, NO 1 (2022)

DOI: http://dx.doi.org/10.23960%2Fkomputasi.v10i1

TABLE OF CONTENTS

ARTICLES

Aplikasi DonasiKu Berbasis Android PDF
Muhamad Bahrul Ulum, Habibullah Akbar, Anik Hanifatul Azizah 1-12
Abstract view : 255 times

Implementasi Metode Naive Bayes pada Sistem Pakar Diagnosis Penyakit Kutu Ikan Gurami (Argunus Indicus) PDF
Agus Wantoro, Heni Sulistiyani, Yodhi Yuniarthe, Arie Setya Putra, Apri Candra Widyawati, Nanda Putra Wicaksono 13-22
Abstract view : 175 times

IMPLEMENTASI FUZZY INFERENCE SYSTEM (FIS) METODE TSUKAMOTO UNTUK MONITORING KUALITAS UDARA PDF
Robby Yuli Endra, Galih Agus Saputra 23-34
Abstract view : 253 times

Implementation of Learning Management System based on Moodle Ruang Belajar for Optimizing Online Learning PDF
Siti Qomariah, Muhammad Nasir, Rahmadani Rahmadani 35-46
Abstract view : 144 times

Perbandingan Nilai K pada Klasifikasi Pneumonia Anak Balita Menggunakan K-Nearest Neighbor PDF
Dwi Kartini, Andi Farmadi, Muliadi muliadi, Dodon Turianto Nugrahadi, Pirjatullah Pirjatullah 47-53
Abstract view : 184 times

Implementasi Reduksi Fitur t-SNE Pada Clustering Gambar Head shape Nematoda PDF
Muhammad Rizky Adriansyah, Mohammad Reza Faisal, Abdul Gafur, Radityo Adi Nugroho, Irwan Budiman, Muliadi Muliadi 54-64
Abstract view : 145 times

Klasifikasi Penyakit Daun Padi menggunakan Random Forest dan Color Histogram PDF
Sarifah Agustiani, Yoseph Tajul Arifin, Agus Junaidi, Siti Khotimatul Wildah, Ali Mustopa 65-74
Abstract view : 318 times

IMPLEMENTASI SUPPORT VECTOR MACHINE (SVM) UNTUK KLASIFIKASI PEDERITA DIABETES MELLITUS PDF
Favorisen R Lumbanraja, Fanni Lufiana, Yunda Heningtyas, Kurnia Muludi 75-83
Abstract view : 253 times

ANALISIS DAN PERANCANGAN KNOWLEDGE MANAGEMENT SYSTEM UNTUK MENINGKATKAN KINERJA PEGAWAI PADA BADAN NARKOTIKA NASIONAL KOTA TANGERANG SELATAN BERBASIS WEBISTE PDF

Rosalina Aprianti, Yuni Sugiarti 84-94

Abstract view : 235 times

Peningkatan Kemenangan Non-Playable Character dalam Permainan Triple Triad Menggunakan Alpha-Beta Pruning PDF

Benedictta Dinda Permatasari, Hanny Haryanto, Erna Zuni Astuti, Erlin Dolphina 95-104

Abstract view : 127 times

P-ISSN

E-ISSN

Download Template Makalah

Sertifikat Akreditasi

Visitors

42,863 93
1,975 43
165 37
127 36
116 33

FLAG counter

45962

Statistik Jurnal

JOURNAL CONTENT

Search

All

Search

Browse

- By Issue
- By Author
- By Title
- Other Journals

Jurnal Komputasi
P-ISSN 2541-0296
E-ISSN 2541-0350

EDITORIAL TEAM

EDITOR-IN-CHIEF

Yunda Heningtyas, [SCOPUS ID: 57225950428] [Sinta ID: 6125066][GS ID: WIi9ZEkAAAAJ][Orchid ID: 0000-0001-5401-9979]; Universitas lampung

MANAGING EDITOR

rizky prabowo, [SCOPUS ID:57201339314][GS:U6JOw0kAAAAJ][OrchidID:0000-0002-9895-9281]Universitas Lampung, Indonesia
Yohana Tri Utami, [SCOPUS ID: 57337180900] [SINTA ID: 6717407] [ORCID ID: 0000-0003-1082-1622] Universitas Lampung, Indonesia

EDITORIAL BOARD

Muhammad Iqbal, [SCOPUS ID: 57211874405]Universitas Negeri Semarang
Anie Rose Irawati, [SINTA ID: 6023917] [SCOPUS ID: 57214646979] Universitas Lampung, Indonesia
Aristoteles ., [SCOPUS ID: 55820317100][orchid:0000-0002-9369-4528][ResearcherID: F-2091-2019] Universitas Lampung
Ossy Dwi Endah Wulansari, Universitas Lampung, Indonesia
Rico Andrian, Universitas Lampung, Indonesia
Purwono Prasetyawan, [SCOPUS ID: 57203960629] Universitas Teknokrat Indonesia, Indonesia
Ardiansyah ., Universitas Lampung, Indonesia
Indra Laksmana, [SCOPUS ID:56192111300] Payakumbuh Agricultural Polytechnic, Indonesia

SERTIFIKAT

Direktorat Jenderal Penguatan Riset dan Pengembangan,
Kementerian Riset, Teknologi, dan Pendidikan Tinggi

RISTEKDIKTI

Kutipan dari Keputusan Direktur Jenderal Penguatan Riset dan Pengembangan, Kementerian Riset, Teknologi, dan Pendidikan Tinggi Republik Indonesia
Nomor: 28/E/KPT/2019
Tentang Hasil Akreditasi Jurnal Ilmiah Periode 5 Tahun 2019

Jurnal Komputasi

E-ISSN: 25410350
Penerbit: Universitas Lampung

Ditetapkan sebagai Jurnal Ilmiah

TERAKREDITASI PERINGKAT 4

Akreditasi berlaku selama 5 (lima) tahun, yaitu
Volume 6 Nomor 1 Tahun 2018 sampai Volume 10 Nomor 1 Tahun 2022
Jakarta, 26 September 2019
Direktur Jenderal Penguatan Riset dan Pengembangan

Dr. Muhammad Dimyati
NIP. 195912171984021001

TERAKREDITASI

Jurnal Komputasi

Vol 10 No. 1, 2022
©2022 Ilmu Komputer Unila Publishing Network all rights reserved

IMPLEMENTASI *SUPPORT VECTOR MACHINE* (SVM) UNTUK KLASIFIKASI PEDERITA *DIABETES MELLIT*US

[1,*)]Favorisen R. Lumbanraja, [2]Fanni Lufiana, [3]Yunda Heningtyas & [4]Kurnia Muludi

[1,2,3 3] Ilmu Komputer, Universitas Lampung, Jl. Prof. Dr. Ir. Sumantri Brojonegoro 35141, Bandar Lampung, Indonesia

Abstrak — *Diabetes Mellitus* (DM) adalah penyakit kronis yang ditandai dengan ketidakmampuan tubuh dalam melakukan metabolisme karbohidrat, lemak, dan protein sehingga terjadinya peningkatan kadar gula darah (*hiperglikemia*) yang disebabkan karena menurunnya kadar hormon insulin. Penyakit diabetes bisa terjadi karena faktor keturunan (genetika) dan pengaruh lingkungan gaya hidup yang tidak sehat. Dalam mengukur kadar gula darah diperlukan tes hemoglobin A1c untuk mendiagnosisi dan mengontrol kondisi penderita diabetes. Adapun tujuan yang dilakukan dalam penelitian ini, yaitu membuat analisa model prediksi untuk klasifikasi penderita diabetes menggunakan R *Shiny* dan mengevaluasi hasil kinerja klasifikasi dari metode *support vector machine*. Terdapat banyak cara dalam mendiagnosa penyakit diabetes, berikut salah satu algoritma *machine learning* yang digunakan dalam kasus klasifikasi pada penelitian ini, yaitu *support vector machine* (SVM). Penelitian ini menggunakan data dari *Diabetes 130-US Hospital For Years* 1999-2008 bersumber dari *UCI Machine Learning Repository* yang terdiri dari 34 variabel dan 84900 *records* dengan teknik pembagian dan pengujian *dataset* menggunakan metode 10-*fold cross validation* serta menggunakan tiga *kernel* dalam melakukan pemodelan menggunakan SVM, yaitu *linear*, *Gaussian*, dan *polynomial*. Hasil penelitian yang didapatkan yaitu sebuah sistem sederhana dari analisa model prediksi untuk klasifikasi penderita diabetes dengan *shiny* yang bertujuan untuk memudahkan pengguna dalam mengetahui hasil prediksi, serta mendapatkan hasil akurasi terbesar yaitu 82,76% dari *kernel gaussian*.

Kata Kunci: *diabetes mellitus*; *HbA1c*; klasifikasi; *support vector machine*; *10-fold cross validation*.

Abstract *— Diabetes Mellitus (DM) is a chronic disease characterized by the body's inability to metabolize carbohydrates, fats, and proteins, resulting in increased blood sugar (hyperglycemia) due to low insulin levels. Diabetes is due to a combination of heredity (genetics) and unhealthy lifestyles. Hemoglobin A1c is a blood test used to diagnose and manage diabetes patients when measuring blood sugar levels. This study aims to analyze predictive models for the classification of people with diabetes using R Shiny and evaluate the results of the support vector machine method's classification performance. There are many ways to diagnose diabetes, and the support vector machine is one of the machine learning algorithms used in this study's classification case (SVM). This study uses data from Diabetes 130-US Hospital For Years 1999-2008, which was sourced from the UCI Machine Learning Repository and consists of 34 variables and 84900 records, with dataset distribution and testing techniques using the 10-fold cross-validation method and three kernels in modeling using SVM, namely linear, Gaussian, and polynomial. The results obtained are a simple predictive model analysis system for classifying people with diabetes with shiny, making it easier for users to find out the prediction results and obtain the highest accuracy result, which is 82.76 percent of the gaussian kernel.*

Keywords: diabetes mellitus; HbA1c; classification; support vector machine; 10-fold cross validation.

** Corresponding author :*
Favorisen R Lumbanraja
Universitas Lampung, Bandar Lampung, Indonesia
favorisen.lumbanraja@fmipa.unila.ac.id

©2022 Ilmu Komputer Unila Publishing Network all rights reserved

1. PENDAHULUAN

Diabetes Mellitus (DM) merupakan jenis penyakit tidak menular yang ditandai dengan meningkatnya kadar glukosa dalam darah akibat metabolisme tubuh yang tidak berfungsi secara normal sehingga organ pankreas tidak mampu menghasilkan hormon insulin sesuai kebutuhan tubuh [1]. Adapun gejala yang dihadapi pada penderita DM seperti polyuria, polydipsia, dan polyphagia [2]. Terdapat 3 jenis penyakit diabetes, yaitu DM tipe 1 (tidak memiliki hormon insulin), DM tipe 2 (hormon insulin tidak dapat bekerja dengan baik sesuai kebutuhan tubuh), dan DM gestasional (dm yang terjadi pada wanita hamil) [3]. HbA1C merupakan glukosa stabil yang terikat pada sel darah merah (hemoglobin) melalui proses non-enzimatis [4]. Pemeriksaan HbA1C penting dilakukan karena sebagai indikator dalam memeriksa tingkat gula darah dalam jangka panjang, diagnosis, dan mengontrol kondisi penderita DM. pemeriksaan HbA1C dilakukan secara berkala setiap 3-6 bulan sekali [5] .

Berdasarkan data dari International Diabetes Federation (IDF) tahun 2019, Indonesia menempati urutan ke-7 dalam jumlah kasus penderita penyakit DM, yaitu sekitar 10.7 juta jiwa, jumlah ini tentunya akan terus bertambah dan diprediksikan pada tahun 2030 akan mencapai angka 13.7 juta jiwa [6]. Berdasarkan hasil laporan riset kesehatan dasar, prevalansi penyakit DM untuk provinsi Lampung sendiri, yaitu sebesar 0.99% dimana Kabupaten Metro memiliki prevalansi penyakit DM paling tinggi, yaitu 2.26% sedangkan Prevalansi penyakit DM paling rendah terdapat di Kabupaten Pesisir Barat, yaitu 0.54% [7].

Beberapa kejadian yang terjadi pada dunia nyata yang berkaitan dengan pengelompokkan atau pelabelan objek dapat dicarikan solusi dengan menerapkan teknik klasifikasi. Sebagai contoh, penelitian pada bidang bio-medikal, teknik klasifikasi dapat digunakan untuk memprediksi penyakit yang diderita oleh pasien sehingga membantu dokter dalam mendiagnosa pasiesn dan memberikan solusi pengobatan yang tepat. Untuk memecahkan masalah klasifikasi, berbagai macam jenis algoritma dalam machine learning dapat diterapkan pada penelitian ini salah satunya, yaitu algoritma SVM (Support Vector Machine) [8]. SVM merupakan algoritma dengan metode supervised learning yang menganalisa data dan mengenali pola [9]. Adapun keunggulan dari algoritma SVM, yaitu dapat digunakan untuk masalah klasifikasi dan regresi dengan kernel linier ataupun non-linier kernel yang dapat menjadi satu kemampuan algoritma pembelajaran untuk klasifikasi serta regresi dengan akurasi yang tinggi[10].

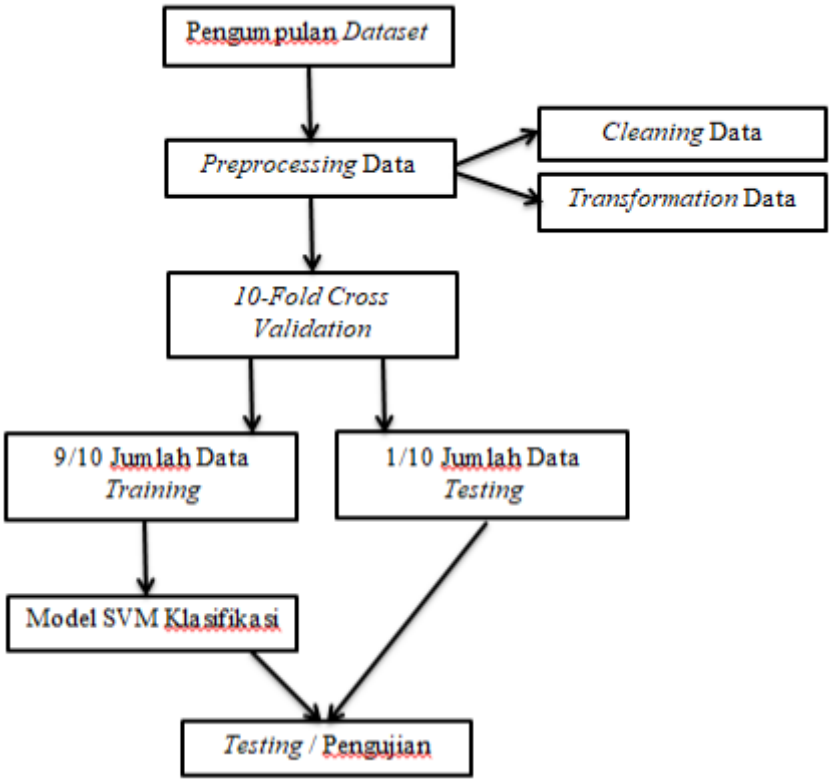
Penelitian yang dilakukan oleh Prasatya, Siregar, & Arianto (2020) meneliti tentang penerapan metode klasifikasi K-Means dan C4.5 untuk memprediksi penyakit diabetes [11]. Penelitian ini melakukan prediksi HbA1C dengan kombinasi metode K-Means dengan C4.5. Data yang digunakan dalam penelitian ini, yaitu ‘Diabetes 130-US Hospitals for Years 1999-2008’ Dataset. Adapun data awal berjumlah 101767 data, namun setelah dilakukan pembersihan data, data berjumlah menjadi 17019 data. Hasil prediksi dengan menggunakan k-fold cross validation didapat sebesar 72%.

Penelitian mengenai penyakit diabetes juga dilakukan oleh Kumari & Chitra (2013) [12]. Penelitian ini bertujuan mengklasifikasikan penyakit diabetes menggunakan metode SVM dengan kernel RBF. Data yang digunakan pada penelitian ini, yaitu data The Pima Indian diabetes dataset dengan data samples yang digunakan, yaitu 460 data. Hasil klasifikasi yang diperoleh mempunyai rata-rata nilai accuracy sebesar 78%, sensitivity 80%, dan specificity 76.5%.

2. METODOLOGI PENELITIAN

Metode yang dilakukan dalam melakukan penelitian ini menggunakan metode SVM dilakukan dengan beberapa tahap yang ditunjukkan pada Gambar 1.

©2022 Ilmu Komputer Unila Publishing Network all rights reserved



Gambar 1. Tahapan Metode Penelitian

Berdasarkan tahapan metode penelitian Gambar 1, berikut penjelasan secara detail dari setiap tahapan yang akan dilakukan pada penelitian ini.

2.1. Pengumpulan Dataset

Tahap awal dalam melakukan penelitian ini, yaitu dengan melakukan pengumpulan data yang berkaitan dengan penelitian. *Dataset* yang digunakan pada penelitian ini adalah data dari *Diabetes 130-US Hospitals For Years 1999-2008 Dataset* yang didapatkan dari situs UCI *Machine Learning Repository*. *Dataset* terdiri dari 101766 *record* dengan 1 label atau kelas dan 34 atribut lainnya yakni *Race*, *Gender*, *Age*, *Time In* Hospital, Diag_1, Diag_2, Diag_3, *Max_glu_serum*, *Metformin*, *Repaglinide*, *Nateglinide*, *Chlorpropamide*, *Glimepiride*, *Acetohexamide*, *Glipizide*, *Glyburide*, *Tolbutamide*, *Pioglitazone*, *Rosiglitazone*, *Acarbose*, *Miglitol*, *Troglitazone*, *Tolazamide*, *Examide*, *Citoglipton*, *Insulin*, *Glyburide-metformin*, *Glipizide-metformin*, *Glimepiride-pioglitazone*, *Metformin-Rosiglitazone*, *Metformin-pioglitazone*, *DiabetesMed*, *Change*, *Readmitted*

2.2. *Preprocessing Data*

Prepocessing data merupakan tahapan sebelum proses pengklasifikasian yang diperlukan untuk membersihkan, menghilangkan, mengubah bentuk data. Tujuannya agar data yang pada proses pengklasifikasiannya [13]. Dalam penelitian ini, *preprocessing data* dilakukan dengan 2 tahap, yaitu dengan cara penghapusan data / *Cleaning Data* dan *Transformation Data* / Data Transformasi (mengubah data dari tipe kategori menjadi tipe *binary*). Setelah data dikumpulkan, *cleaning data* dilakukan dengan menghapus data yang tidak sesuai atau tidak diperlukan dalam penelitian, seperti menghapus data kosong yang tidak memiliki nilai pada data penelitian dan mengambil data penting yang akan diteliti secara manual. Setelah itu, melakukan transformasi data dengan cara mengubah data dari *categorical* menjadi *binary*. Jumlah data yang dihasilkan pada tahap *cleaning data* menjadi 84900 data dan jumlah atribut / variabel yang dihasilkan pada tahap *transformation data* menjadi 85 atribut / variabel.

©2022 Ilmu Komputer Unila Publishing Network all rights reserved

2.3. *K-Fold Cross Validation*

Tahap selanjutnya, yaitu melakukan *k-fold cross validation*, tahapan ini dilakukan untuk mengetahui kinerja klasifikasi dengan memotong data menjadi fold-fold data uji dan data latih [14]. Proses *k-fold cross validation* diawali dengan membagi data sejumlah *n-fold* yang diinginkan dengan ukuran yang sama, untuk penelitian ini menggunakan *10-fold cross validation* karena penggunaan jumlah *fold* terbaik untuk uji validitas [15]. Selanjutnya proses membagi data penelitian menjadi data *testing* dan data *training* yang dilakukan sebanyak n kali. Untuk pembagian data meliputi 1/10 jumlah data sebagai data *testing*, dan 9/10 jumlah data sebagai data *training*. Ulangi proses sampai setiap iterasi n yang berfungsi sebagai set uji dan mengambil rata-rata skor yang tercatat. Rata-rata skor yang tercatat akan menjadi metrik performa untuk model.

2.4. Pemodelan SVM Klasifikasi

Dalam penelitian ini, algoritma yang digunakan adalah algoritma *Support Vector Machine* untuk melakukan proses klasifikasi. Klasifikasi merupakan proses untuk menghasilkan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk memperkirakan kelas yang tidak diketahui dari suatu objek [16]. Untuk tahap pemodelan svm dilakukan dengan tujuan untuk menemukan *hyperplane* terbaik yang memisahkan dua buah *class* pada *input space* [17]. Tingkat akurasi pada model yang akan dihasilkan dalam proses peralihan dengan SVM sangat bergantung terhadap fungsi *kernel* dan parameter yang digunakan . Adapun dalam tahap ini, fungsi *kernel* yang digunakan yaitu *kernel linear, Gaussian,* dan *polynomial*. Beberapa tipe kernel yang digunakan dapat dijabarkan pada Tabel 1.

Tabel 1. Tipe *Kernel*

Nama Kernel	Fungsi Kernel
Kernel Linear	$K(x,y) = (x^T y)$
Kernel Gaussian	$K(x,y) = exp\left\{-\frac{\lvert x-y\rvert^2}{2\gamma^2}\right\}$
Kernel Polynomial	$K(x,y) = (\gamma x^T y + coef.0)^d$

Selanjutnya, data *training* atau data latih digunakan untuk melakukan proses pelatihan. Proses *training* menggunakan *Multi-Class* SVM, dengan k buah model SVM biner (k adalah jumlah kelas), dalam penelitian ini jumlah kelas yang digunakan, yaitu 4 kelas.

2.5. *Testing* / Pengujian

Tahap terakhir pada penelitian ini adalah dengan melakukan *testing* atau pengujian. Model SVM yang telah didapatkan pada tahapan sebelumnya digunakan untuk menguji data *testing* dari proses pembagian data penelitian. Pengujian dalam penelitian ini menggunakan *confusion matrix* untuk mendapatkan nilai *accuracy*, *sensitivity*, dan *specificity*. Berikut tabel *confusion matrix* dapat dilihat pada Tabel 2 dan nilai pengujian yang dilakukan menggunakan *confusion matrix* [18].

©2022 Ilmu Komputer Unila Publishing Network all rights reserved

Tabel 2. Tabel *Confusion Matrix*

Correct Classification	***Classified as***	
	Actual "+"	***Actual "-"***
Predicted "+"	TP (*True Positive*)	FP (*False Positive*)
Predicted "-"	FN (*False Negative*)	TN (*True Negative*)

3. HASIL DAN PEMBAHASAN

3.1. *Preprocessing*

Sebelum melakukan percobaan model klasifikasi pada penelitian ini, perlu dilakukan proses *preprocessing* terlebih dahulu dengan langkah-langkah sebagai berikut.

3.1.1 *Cleaning Data* (Pembersihan / Penghapusan Data)

Melakukan pembersihan data yang tidak sesuai / tidak dapat dibaca dan tidak dapat diolah pada saat melakukan percobaan klasifikasi. Dalam penelitian ini data awal berjumlah 101766, setelah dilakukan proses pembersihan data, data menjadi 84900 data. Berikut Tabel 3 merupakan jumlah data setelah dilakukan proses pembersihan data.

Tabel 3. Perubahan Jumlah Data Setelah Proses Pembersihan Data.

Data	Data Sebelum Dibersihkan	Data Setelah Dibersihkan
Norm (DM kontrol baik)	4990	4258
>7 (DM kontrol sedang)	3812	3335
>8 (DM kontrol buruk)	8216	6850
None	84748	70457
Jumlah	101766	84900

Berdasarkan Tabel 3 jumlah data awal pada data DM kontrol baik sebanyak 4990 data, DM kontrol sedang sebanyak 3812 data, DM kontrol buruk sebanyak 8216 data, dan *None* sebanyak 84748 data. Setelah dibersihkan terdapat pengurangan jumlah data sebanyak 732 data pada data DM kontrol baik, sehingga jumlah data menjadi 4258 data. Pengurangan jumlah data DM kontrol sedang sebanyak 477 data, sehingga jumlah data menjadi 3335 data. Pengurangan jumlah data DM kontrol buruk sebanyak 1366 data, sehingga jumlah data menjadi 6850 data. Pengurangan jumlah data *None* sebanyak 14291 data, sehingga jumlah data menjadi 70457 data. Sehingga data pada penelitian ini memiliki *imbalance data* (data tidak seimbang) antara kelas negatif (*None)* dengan kelas lainnya yang positif (DM terkontrol baik, DM terkontrol sedang, dan DM terkontrol buruk). Data penelitian pada kelas negatif memiliki jumlah lebih besar, yaitu 70457 data dari kelas positif yang berjumlah 14443 data.

3.1.2 *Transformation Data*

Transformasi Data dilakukan supaya dapat memperoleh data yang lebih berkualitas dengan cara mengubah tipe data *categorical* menjadi tipe data *binary*. sebelumnya terdapat 29 atribut bertipe data *categorical*, setelah dilakukan proses *tranformation data* jumlah atribut menjadi 85 atribut yang akan digunakan untuk penelitian ini. Berikut Tabel 4 merupakan hasil dari *transformation data*.

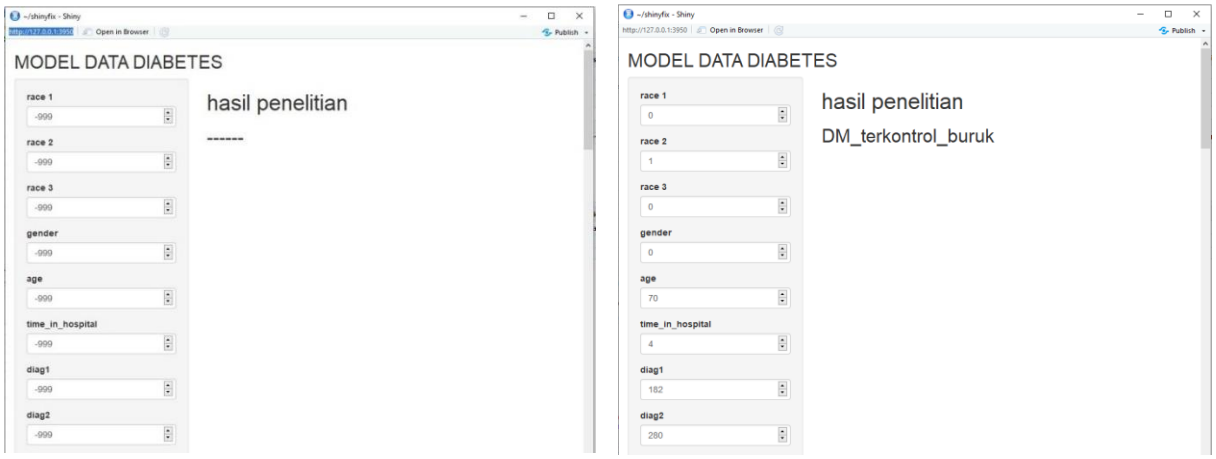
©2022 Ilmu Komputer Unila Publishing Network all rights reserved

Tabel 4. Data Hasil *Transformation Data*

Insulin		insulin1	insulin2	insulin3	insulin4
Up		0	1	0	0
Steady	→	1	0	0	0
Down		0	0	1	0
No		0	0	0	1

3.2. Model Hasil Prediksi

Penelitian ini membuat contoh model hasil prediksi dengan metode *support vector machine* untuk klasifikasi penderita *diabetes* menggunakan R *Shiny*. *Shiny* yaitu sebuah paket dalam R yang mengijinkan penggunanya membangun web apps yang interaktif menggunakan kode R [19]. Tujuan dibuatnya model ini agar hasil prediksi yang telah diperoleh bisa di implementasikan ke dalam sebuah sistem yang dapat memudahkan pengguna untuk mengetahui hasil prediksi dari penderita *diabetes*. Berikut tampilan awal dari sistem terdapat pada Gambar 2 yang merupakan tampilan awal sistem untuk memasukan nilai dari setiap atribut yang ada di sistem supaya sistem dapat mengeluarkan hasil analisis prediksi tersebut, serta Gambar 3 menampilkan hasil analisa model prediksi menggunakan metode *support vector machine* dengan *Shiny*.



a b

Gambar 3. *Interface* (a) Awal Sistem Sebelum Data Di *Input*; (b) Hasil Analisa Model Prediksi Dengan *Shiny* Setelah Data Di *Input* Oleh *User* Sesuai Nilai Atribut

3.3. Percobaan Hasil Kinerja Dari Klasifikasi SVM

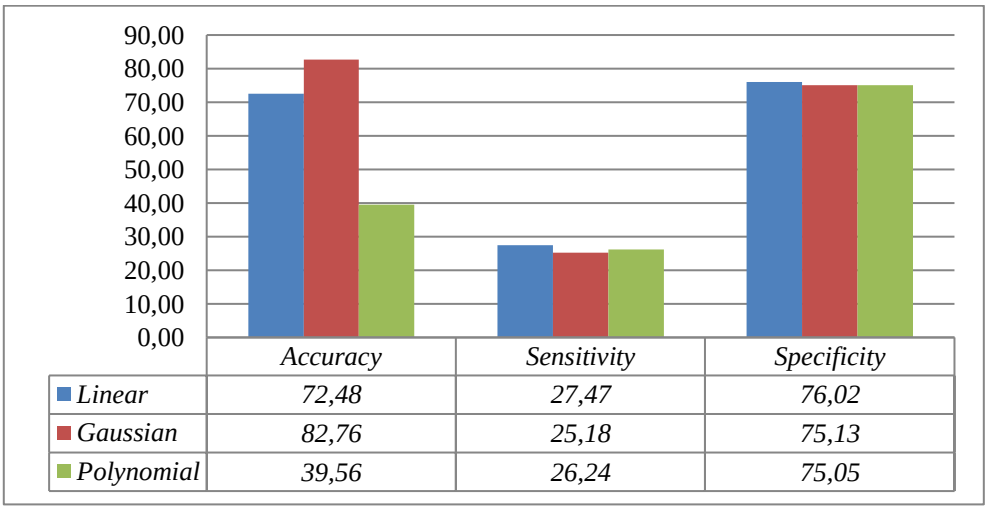
Dalam melakukan proses percobaan mengukur hasil kinerja suatu model klasifikasi dilakukan dengan menggunakan metode *confusion matrix* [18], penelitian ini menggunakan 84900 data dengan 85 atribut yang dikelompokan kedalam 4 kelas, yaitu kelas *Norm* (DM kontrol baik), *>7* (DM kontrol sedang), *>8* (DM kontrol buruk) dan *None*. Penelitian menggunakan 3 *kernel* dari model algoritma SVM, yaitu *Kernel Linear*, *Kernel Gaussian*, dan *Kernel Polynomial*. Pengujian pada penelitian ini menggunakan 10-*fold cross validation* dengan membagi data menjadi 2 bagian, yaitu data *testing* berjumlah 8490 data dan data *training* berjumlah 76410 data. Didapatkan hasil kinerja klasifikasi menggunakan persamaan 1, 2, dan 3 dari masing-masing *kernel* pada Tabel 5.

©2022 Ilmu Komputer Unila Publishing Network all rights reserved

Tabel 5. Hasil Kinerja Klasifikasi Dari Setiap *Kernel*

	Accuracy	*Sensitivity*	*Specificity*
Kernel Linear	72.48%	27.47%	76.02%
Kernel Gaussian	82.76%	25.18%	75.13%
Kernel Polynomial	39.56%	26.24%	75.05%

3.4. Pembahasan



Gambar 4. Grafik Hasil Prediksi Menggunakan *Kernel Linear, Gaussian, dan Polynomial*

Berdasarkan Gambar 4 dapat dilihat bahwa hasil akurasi terbesar terdapat pada *kernel Gaussian*, yaitu 82,76%. Hasil *sensitivity* bernilai kecil karena data penelitian ini memiliki *imbalance data* (data tidak seimbang) dimana kasus jumlah pasien yang mengklasifikasikan positif menderita diabetes lebih kecil dibanding dengan jumlah data pasien penderita diabetes yang mengklasifikasikan negatif pada hasil prediksi sehingga hasil *sensitivity* tidak bagus. Sedangkan nilai *specificity* tinggi karena pada data penelitian ini menunjukan bahwa jumlah pasien penderita diabetes lebih banyak bernilai negatif yang artinya pasien yang benar-benar negatif diabetes di identifikasi benar bahwa pasien tersebut memang tidak memiliki penyakit diabetes.

4. KESIMPULAN

Berdasarkan hasil pengujian dari implementasi model klasifikasi SVM, dapat disimpulkan bahwa model prediksi untuk klasifikasi penderita diabetes dapat diimplementasikan kedalam sebuah sistem yang dapat digunakan dalam penanganan kejadian pada penyakit diabetes. Hasil analisis korelasi pada penelitian ini menunjukkan bahwa setengah variabel dari *dataset* tidak memiliki korelasi (hubungan). Serta hasil klasifikasi menggunakan metode *support vector machine* (SVM) didapatkan hasil rata-rata *accuracy* terbesar diperoleh dari *kernel Gaussian*, yaitu sebesar 82,76%, hasil rata-rata *sensitivity* sebesar 27.47% dan hasil *specificity* sebesar 76.02%. Saran untuk penelitian selanjutnya, diperlukannya pengolahan data yang bisa menangani jumlah kelas yang *imbalance* sehingga hasil kinerja dari *sensitivity* dan *specificity* bisa lebih baik lagi.

©2022 Ilmu Komputer Unila Publishing Network all rights reserved

DAFTAR PUSTAKA

[1] R. D. Gradianta and I. , "Klasifikasi Pasien Penderita Diabetes Mellitus Tipe 2 Menggunakan Metode Analisis Diskriminan Hybrid Algoritma Genetika," *JURNAL SAINS DAN SENI POMITS* , vol. 3, 2014.

[2] A. A. Aljumah, M. G. Ahamad and M. K. Siddiqui, "application of data mining diabetes health care in young and old patients," vol. 25, pp. 127-136, 2012.

[3] American Diabetes Association, "Position statement: Standards of Medical Care in Diabetes 2010," *Diab Care,* 2010.

[4] E. Afrianti, N. Bahri, Y. and S. Elmatris, "EFEK HIDROTERAPI PADA PENURUNAN KADAR GULA DARAH SESAAT (KGDS) TERHADAP PENDERITA DIABETES MELITUS TIPE 2," *Majalah Kedokteran Andalas,* vol. 36, 2012.

[5] A. Que, I. W. P. S. Yasa and A. W. Lestari, "Gambaran Hasil Pemeriksaan Kadar HbA1C pada Penderita Diabetes Mellitus di Laboratorium Rumah Sakit Umum Surya Husada Tahun 2013," 2013.

[6] International Diabetes Federation, "Global and Regional Diabetes Prevalence Estimates For 2019 and Projections For 2030 and 2045: Results From The International Diabetes Federation Diabetes Atlas, 9th Edition," *Diabetes Research and Clinical Practice,* 2019.

[7] Tim Riskesdas, Laporan Nasional RISKESDAS 2018, Jakarta: Lembaga Penerbit Badan Penelitian dan Pengembangan Kesehatan (LPB), 2019.

[8] F. Yunita, "SISTEM KLASIFIKASI PENYAKIT DIABETES MELLITUS MENGGUNAKAN METODE K-NEAREST NEIGHBOR (K-NN)," *Jurnal BAPPEDA,* vol. 2, April 2016.

[9] H. Bhavsar and M. H. Panchal, "A Review On Support Vector Machine For Data Classification," *International Journal of Advanced Research In Computer Engineering and Technology,* pp. 2278-1328, 2012.

[10] O. Maimon and L. Rokach, "Data mining and knowledge discovery handbook," 2010.

[11] A. Prasatya, R. R. A. Siregar and R. Arianto, "Penerapan Metode K-Means Dan C4.5 Untuk Prediksi Penderita Diabetes," *Pengkajian dan Penerapan Teknik Informatika,* vol. 13, Maret 2020.

[12] V. A. Kumari and R. Chitra, "Classification Of Diabetes Disease Using Support Vector Machine," *International Journal of Engineering Research and Applications,* vol. 3, no. 2, 2013.

[13] F. A. Muttaqin and A. M. Bachtiar, "IMPLEMENTASI TEKS MINING PADA APLIKASI PENGAWASAN PENGGUNAAN INTERNET ANAK "DODO KIDS BROWSER"," *Jurnal Ilmiah Komputer dan Informatika (KOMPUTA),* 2019.

[14] E. Pandie, "Implementasi Algoritma Data Mining K Nearest Neighbour (KNN) Dalam Pengambilan Keputusan Pengajuan Kredit," 2012.

[15] Nurhayati, I. K. Hadihardaja, I. Soekarno and M. Cahyono, "A Study of Hold-Out and K-Fold Cross Validation for Accuracy of Groundwater Modeling in Tidal Lowland Reclamation Using Extreme Learning Machine," *IEEE,* 2014.

[16] I. F. Yuliati, S. Wulandary and P. . R. Sihombing, "Penerapan Metode Support Vector Machine (SVM) dan Backpropagation Neural Network (BPNN) dalam Pengklasifikasian Pasangan Usia Subur di Jawa Barat," *Jurnal Statistika dan Aplikasinya,* vol. 4, no. 1, 2020.

©2022 Ilmu Komputer Unila Publishing Network all rights reserved

[17] A. S. Ritonga and E. S. Purwaningsih, "PENERAPAN METODE SUPPORT VECTOR MACHINE (SVM) DALAM KLASIFIKASI KUALITAS PENGELASAN SMAW (SHIELD METAL ARC WELDING)," *Jurnal Ilmiah Edutic,* vol. 5, 2018.

[18] M. F. Rahman, M. I. Darmawidjadja and D. Alamsyah, "KLASIFIKASI UNTUK DIAGNOSA DIABETES MENGGUNAKAN METODE BAYESIAN REGULARIZATION NEURAL NETWORK (RBNN)," *Jurnal Informatika,* vol. 11, Januari 2017.

[19] W. Chang, "Web Application Framework for R," *Package 'shiny',* 2021.