

## Klasifikasi Diabetes Menggunakan Algoritma *Support Vector Machine* Radial Basis Function

Muhammad Hilmy Haidar Aly

UPN “Veteran” Jawa Timur

Alamat: Jl. Rungkut Madya No.1, Surabaya

**Abstract.** *Diabetes mellitus (DM) is a chronic disease associated with high levels of sugar or glucose in the blood caused by pancreatic and insulin disorders. According to data from the Ministry of Health of the Republic of Indonesia, Diabetes is the third leading cause of death in Indonesia with a percentage of 6.7%. The high rate prompted this study to conduct early detection. One approach that has been widely used is the use of the Support Vector Machine algorithm in predictive modeling. The method was chosen because it was proven in previous studies to get quite high accuracy. Several preprocessing methods were performed to prepare the data for the classification process. The data used involved parameters such as pregnancy, glucose, blood pressure, skin thickness, insulin, weight, heredity, and age. Based on experimental results in ongoing system testing, the maximum performance result is 87% using SVM.*

**Keywords:** *Classification, Diabetes, And Support Vector Machine.*

**Abstrak.** Diabetes melitus (DM) merupakan penyakit kronis yang berhubungan dengan tingginya kadar gula atau glukosa dalam darah yang diakibatkan oleh gangguan pankreas dan insulin. Menurut data dari Kementerian Kesehatan Republik Indonesia, Diabetes merupakan penyebab kematian terbesar nomor 3 di Indonesia dengan presentase sebesar 6.7%. Tingginya angka tersebut mendorong penelitian ini untuk melakukan deteksi dini. Salah satu pendekatan yang telah digunakan secara luas adalah penggunaan algoritma *Support Vector Machine* dalam pemodelan prediktif. Metode tersebut dipilih karena terbukti pada penelitian sebelumnya mendapatkan akurasi yang cukup tinggi. Beberapa metode preprocessing dilakukan untuk mempersiapkan data agar dapat dilakukan proses klasifikasi. Data yang digunakan melibatkan parameter seperti kehamilan, glukosa, tekanan darah, ketebalan kulit, insulin, berat badan, keturunan, dan umur. Berdasarkan hasil eksperimen dalam pengujian sistem yang berlangsung diperoleh hasil performansi maksimum sebesar 84% dengan menggunakan SVM.

**Kata kunci:** Klasifikasi, Diabetes, Dan *Support Vector Machine*.

### LATAR BELAKANG

Diabetes merupakan salah satu masalah kesehatan utama di seluruh dunia. Ini adalah suatu kondisi ketika tubuh tidak cukup untuk menghasilkan atau menggunakan hormon insulin yang membawa glukosa ke dalam sel-sel tubuh dan memungkinkan glukosa untuk masuk dan menjadi bahan bakar mereka. Diabetes mellitus umumnya disebut sebagai “diabetes” yaitu penyakit kronis yang berhubungan dengan tingginya kadar glukosa (gula) dalam darah (Devi, Bai, & Nagarajan, 2019).

Penyakit diabetes ini merupakan penyakit yang cukup berbahaya. Komplikasi jangka panjang diabetes berkembang secara bertahap. Semakin lama orang menderita diabetes dan semakin tidak terkontrolnya gula darahnya, semakin tinggi risiko komplikasinya. Pada akhirnya, komplikasi diabetes mungkin dapat menyebabkan kelumpuhan atau bahkan

mengancam jiwa. Kemungkinan komplikasi biasanya yang terjadi meliputi penyakit kardiovaskular, kerusakan pada saraf, ginjal, mata, kaki, kulit, pendengaran, dan bahkan dapat menyebabkan depresi (Mayo Clinic, 2018).

Pada tahun 2019, International Diabetes Federation mempublikasikan penelitiannya mengenai jumlah penderita penyakit diabetes di seluruh dunia. Indonesia menempati peringkat kedua setelah China di wilayah pasifik barat dengan penderita 10,7 juta orang dari total populasi dewasa 172,2 juta. Itu artinya ada sekitar 6,2% orang dewasa di Indonesia menderita penyakit diabetes.

Untuk mengendalikan peningkatan diabetes, diperlukan pendeteksian dini terhadap penyakit ini. Adapun cara untuk melakukan pendeteksian dini penyakit diabetes yaitu dengan menggunakan suatu sistem. Pendeteksian menggunakan sistem salah satunya diawali dengan mengolah data terkait faktor - faktor yang dapat mempengaruhi seseorang menderita diabetes seperti (1) jumlah kali hamil; (2) konsentrasi glukosa plasma 2 jam dalam tes toleransi glukosa oral; (3) tekanan darah diastolik (mm Hg); (4) triceps ketebalan lipatan kulit (mm); (5) serum insulin 2-Jam ( $\mu\text{U} / \text{ml}$ ); (6) indeks massa tubuh (berat badan dalam kg / (tinggi dalam m)<sup>2</sup>); (7) fungsi diabetes silsilah; dan umur (tahun). Selanjutnya data tersebut diolah menggunakan metode klasifikasi. Saat ini metode klasifikasi yang cukup populer adalah Support Vector Machine (SVM).

Tujuan utama dari SVM adalah membangun sebuah model (yang dihasilkan dari data pelatihan) yang dapat memprediksi data tes yang diberikan. SVM saat ini telah berhasil digunakan untuk menyelesaikan permasalahan pengenalan pola, deteksi penyakit, deteksi wajah, penarikan informasi, dan lain sebagainya. Salah satu keuntungan mendeteksi diabetes menggunakan sistem dengan metode SVM adalah waktu yang diperlukan singkat.

Penelitian mengenai penyakit diabetes dilakukan oleh (Sisodia dan Sisodia, 2018) yang bertujuan untuk merancang model yang dapat memprediksi kemungkinan terjadinya diabetes pada pasien. Metode yang digunakan sebagai klasifikasi adalah Decision Tree, SVM, dan Naive Bayes. Hasil yang diperoleh menunjukkan akurasi masing-masing algoritma adalah Naïve Bayes 76,30%, Support Vector Machine 65,10%, Decision Tree 73,82%.

Oleh karena itu, berdasarkan latar belakang dan hasil penelitian sebelumnya, penelitian ini bertujuan utama untuk mengklasifikasi penyakit diabetes dengan menggunakan metode SVM serta RBF kernel. Tugas utama yang terkait dengan penelitian ini adalah pengumpulan data, pra-pemrosesan data, dan pengembangan model klasifikasi. Dataset yang menjadi patokan terkait penelitian diabetes ini diperoleh dari situs web Kaggle.

## KAJIAN TEORITIS

### 1. Machine Learning

Machine learning merupakan serangkaian teknik yang dapat membantu dalam menangani dan memprediksi data yang sangat besar dengan cara mempresentasikan data-data tersebut dengan algoritma pembelajaran (Danukusumo, 2017).

Istilah machine learning sendiri pertama kali didefinisikan oleh Arthur Samuel pada tahun 1959. Menurut Arthur Samuel, machine learning adalah suatu bidang ilmu komputer yang memberikan kemampuan pembelajaran kepada komputer untuk mengetahui sesuatu tanpa pemrograman yang jelas.

### 2. Data Pre-processing

Pre-processing data atau pembersihan data merupakan tahapan awal yang dilakukan sebelum masuk ke tahap pembentukan model machine learning. Pembersihan data berisi berbagai proses yang bertujuan untuk melakukan perbaikan pada data yang akan diteliti. Melakukan pembersihan data adalah salah satu hal yang penting, karena data yang masih mentah cenderung tidak siap dan perlu dikaji sedemikian rupa terlebih dahulu agar sesuai dengan kebutuhan sistem. Pada pembersihan data ini, kasus yang sering dijumpai adalah hilangnya sebagian data atau missing values. Maksud dari missing values adalah adanya beberapa fitur yang tidak memiliki nilai informasi atau kosong yang dimana seharusnya memiliki nilai.

### 3. Data Training dan Testing

Menurut Prasetyo (2014), data untuk pengujian klasifikasi dibagi menjadi data training dan data testing. Data atau vektor yang sudah diketahui sebelumnya untuk label kelompok dan digunakan untuk membangun model classifier disebut data training. Data atau vektor yang belum diketahui (dianggap belum diketahui) label kelompoknya sehingga menggunakan model *classifier* yang sudah pernah dibangun untuk diketahui label kelompoknya disebut data testing.

Data training digunakan oleh algoritma klasifikasi untuk membentuk sebuah model classifier. Model ini merupakan representasi pengetahuan yang akan digunakan untuk memprediksi kelompok dari data baru yang belum pernah ada. Data testing digunakan untuk mengukur sejauh mana *classifier* berhasil melakukan klasifikasi dengan benar. Data yang ada pada data testing seharusnya tidak boleh ada pada data training sehingga dapat

diketahui apakah model *classifier* sudah tepat atau belum dalam melakukan klasifikasi (Witten dan Eibe, 2011).

#### 4. Algoritma SVM

Algoritma Support Vector Machine (SVM) ditemukan oleh Vladimir N. Vapnik dan Alexey Ya. Chervonenkis pada tahun 1963. Pada tahun 1963. Pada tahun 1992, Bernhard E. Boser, Isabelle M. Guyon dan Vladimir N. Vapnik mengusulkan cara untuk membuat nonlinier *classifier* dengan menerapkan trik kernel pada maximum-margin hyperplanes (Boser, Guyon, & Vapnik, 1992). Kemudian konsep standar SVM saat ini diajukan Corinna Cortes dan Vapnik pada tahun 1993 dan diterbitkan pada tahun 1995 (Cortes & Vapnik, 1995).

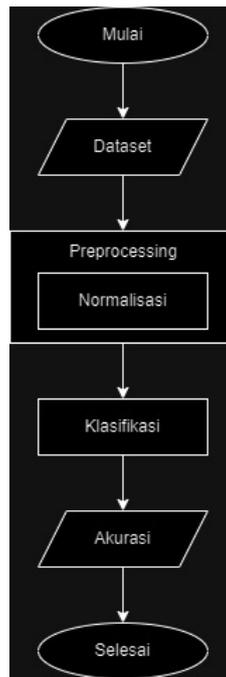
Prinsip yang mendasar dari SVM adalah bagaimana mencari fungsi *hyperplane* (garis pemisah) yang dapat memisahkan antara kedua kelas secara maksimal. Maksimal yang dimaksud adalah *hyperplane* dapat memisahkan data kedua kelas dengan margin yang paling baik. Margin merupakan jarak garis *hyperplane* dengan anggota – anggota terdekat dari kedua kelas. Margin yang mampu memisahkan kelas secara maksimal disebut *Optimal Hyperplane*.

#### 5. Radial Basis Function (RBF) Kernel

RBF kernel merupakan fungsi kernel yang biasa digunakan dalam analisis ketika data tidak terpisah secara linear. RBF kernel memiliki dua parameter yaitu *Gamma* dan *Cost*. Parameter *Cost* atau biasa disebut sebagai C merupakan parameter yang bekerja sebagai pengoptimalan SVM untuk menghindari misklasifikasi di setiap sampel dalam training dataset. Parameter *gamma* menentukan seberapa jauh pengaruh dari satu sampel training dataset dengan nilai rendah berarti “jauh”, dan nilai tinggi berarti “dekat”. Dengan *gamma* yang rendah, titik yang berada jauh dari garis pemisah yang masuk akal dipertimbangkan dalam perhitungan untuk garis pemisah. Ketika *gamma* tinggi berarti titik-titik berada di sekitar garis yang masuk akal akan dalam perhitungan (Patel, 2017).

### **METODE PENELITIAN**

Sistem yang dibangun pada penelitian ini terdiri dari beberapa proses. Adapun bagan metode penelitian yang dibangun ditunjukkan oleh Gambar 1.



**Gambar 1.** Desain Metode Penelitian

Penelitian ini mengusulkan rancangan sistem yang dapat melakukan klasifikasi terhadap data diabetes. Dataset yang digunakan melibatkan atribut yang telah dijelaskan pada bagian sebelumnya, pada Gambar 1 menunjukkan alur penelitian yang diusulkan yang terdiri dari beberapa modul yakni *Preprocessing*, data split untuk membagi data menjadi data training dan testing, dan klasifikasi menggunakan algoritma SVM.

### Dataset

Dalam penelitian ini akan menggunakan dataset diabetes yang didapatkan dari situs website Kaggle. Atribut - atribut yang ada didalamnya adalah atribut yang digunakan pada proses yang biasa dilakukan laboratorium untuk mendeteksi pasien diabetes atau tidak. Pada dataset ini terdapat 769 record (pasien).

**Tabel 1.** Tipe Dataset

Atribut	Tipe Data	Karakteristik
Kehamilan (X1)	Numerik	Atribut
Glukosa (X2)	Numerik	Atribut
Tekanan Darah (X3)	Numerik	Atribut
Ketebalan Kulit (X4)	Numerik	Atribut
Insulin (X5)	Numerik	Atribut
Berat badan (X6)	Numerik	Atribut
Keturunan (X7)	Numerik	Atribut
Umur (X8)	Numerik	Atribut
Hasil (Y)	Biner(Ya/Tidak)	Kelas

### ***Preprocessing***

Preprocessing merupakan sebuah langkah penting dalam proses penambangan data. Data yang akan digunakan dalam proses penambangan data tidak selalu dalam kondisi terbaik untuk diproses. Ada kalanya dalam data tersebut terdapat beberapa masalah yang nantinya dapat mempengaruhi hasil yang diberikan dari proses penambangan itu sendiri seperti terdapat nilai yang hilang, data yang berlebihan, outliner, atau format data yang tidak sesuai dengan sistem. Oleh karena itu untuk mengatasi masalah tersebut perlu dilakukan tahap preprocessing. Preprocessing merupakan salah satu tahapan dalam menghilangkan masalah yang dapat mengganggu hasil dari pada proses klasifikasi data. Salah satu tahapan yang dapat dilakukan yaitu Normalisasi data.

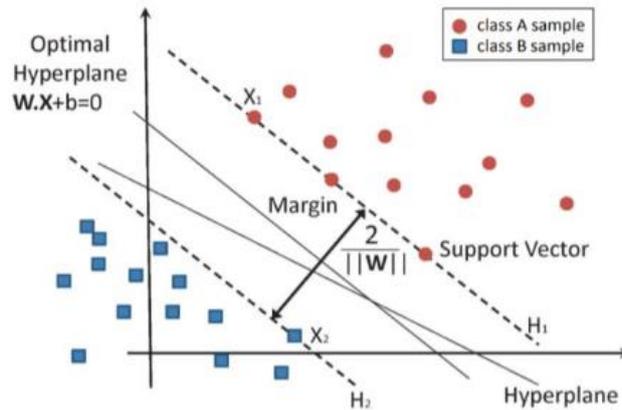
Normalisasi data merupakan proses dilakukan transformasi sebuah atribut numerik yang diskalkan ke dalam sebuah bentuk lebih sederhana seperti 0 dan 1. Untuk melakukan hal tersebut dapat dilakukan dengan cara menggunakan *Min-Max Normalization*. Dalam metode tersebut data akan ditransformasikan secara linear dari suatu nilai menjadi nilai baru lainnya[2], rumusnya sebagai berikut.

$$v' = \frac{v - \min_A}{\max_A - \min_A} (\text{new\_max}_A - \text{new\_min}_A) + \text{new\_min}_A$$

### ***Classification***

Tahapan klasifikasi adalah tahapan dilakukannya prediksi terhadap dataset untuk menemukan suatu pola yang sebelumnya telah melewati tahap *preprocessing* dan *Data Split*. Pada tahap ini metode klasifikasi yang digunakan adalah algoritma *Support Vector Machine*.

Prinsip yang mendasar dari SVM adalah bagaimana mencari fungsi *hyperplane* (garis pemisah) yang dapat memisahkan antara kedua kelas secara maksimal. Maksimal yang dimaksud yaitu *hyperplane* dapat memisahkan data kedua kelas dengan margin yang paling bagus. Margin merupakan jarak garis *hyperplane* dengan anggota terdekat dari kedua kelas. Margin yang mampu memisahkan kelas secara maksimal disebut *Optimal Hyperplane*.



**Gambar 2.** Klasifikasi data menggunakan SVM [3]

Pada gambar tersebut menunjukkan bahwa garis  $H_1$ ,  $H_2$ , dan hyperplane merupakan pemisah kedua kelas. Dimana  $X$  adalah dot product dari variabel dan konstanta pada setiap notasi dan  $W$  adalah nilai yang tegak lurus dengan  $X$ .

$$w \cdot X_i + b \leq -1 \quad (1)$$

Persamaan 1 merupakan hyperplane yang bersinggungan terhadap data yang ada pada kelas A ( $H_1$ ).

$$w \cdot X_i + b \geq +1 \quad (2)$$

Persamaan 2 merupakan hyperplane yang bersinggungan terhadap data yang ada pada kelas B ( $H_2$ ).

$$w \cdot X + b = 0 \quad (3)$$

Dan Persamaan 3 merupakan hyperplane yang berada di antara hyperplane kelas A dan kelas B (Garis Hyperplane). Sedangkan untuk data yang bersinggungan dengan  $H_1$  di kelas A dan  $H_2$  di kelas B disebut dengan Support Vector.

Pencarian titik minimal disebut juga dengan Quadratic Programming (QP). Penentuan margin diperlukan untuk menentukan titik minimal yaitu dengan  $\frac{1}{\|w\|}$ . Berikut ini adalah persamaan untuk mencari titik minimal [4].

$$\min_w \tau(w) = \frac{1}{2} \|w\|^2 \quad (4)$$

Dengan memerhatikan nilai constrain:

$$y_i(X_i \cdot w + b) - 1 \geq 0, \quad \forall_i \quad (5)$$

Sampel data yang ada tidak semua memiliki data yang terpisah secara linear sehingga tidak bisa menggunakan SVM *linear*. Apabila dipaksakan nantinya akan memberikan hasil klasifikasi yang buruk dan tidak optimal. Sehingga harus merubah SVM linear menjadi SVM non-linear agar dapat berjalan dengan optimal, caranya yaitu dengan memanfaatkan metode kernel. Pendekatan ini berbeda dengan metode klasifikasi secara umum, yang sebenarnya mengurangi dimensi awal untuk menyederhanakan proses komputasi dan meningkatkan performansi[5].

### Akurasi

Tingkat akurasi sebuah classifier untuk data testing adalah rasio perbandingan jumlah data testing yang dapat diklasifikasikan dengan benar dengan jumlah seluruh data. Akurasi dapat dinyatakan dalam satuan persen dengan mengalikan hasil perhitungan dengan 100%.

## HASIL DAN PEMBAHASAN

Pada penelitian ini terdapat 3 macam scenario pengujian yang dilakukan. Pengujian pada skenario pertama menggunakan random state 42. Pengujian pada skenario kedua menggunakan random state 50. Kemudian untuk pengujian ketiga menggunakan random state 60. Setiap skenario tersebut terbagi lagi menjadi 3 pembagian data training dan testing yakni 80 dan 20%, selanjutnya 70 dan 30%, dan yang terakhir 90 dan 10%.

### Pengujian Skenario Pertama

Pada skenario pengujian pertama ini dilakukan untuk mengetahui tingkat akurasi yang dihasilkan menggunakan random state 42 menggunakan algoritma SVM. Setelah dilakukan pengujian di dapatkan hasil sebagai berikut.

**Tabel 2.** Skenario Pertama

Pembagian Data	Akurasi
80 dan 20	73%
70 dan 30	82%
90 dan 10	87%

### Pengujian Skenario kedua

Pada skenario pengujian pertama ini dilakukan untuk mengetahui tingkat akurasi yang dihasilkan menggunakan random state 50 menggunakan algoritma SVM. Setelah dilakukan pengujian di dapatkan hasil sebagai berikut.

**Tabel 3. Skenario Kedua**

<b>Pembagian Data</b>	<b>Akurasi</b>
80 dan 20	78%
70 dan 30	76%
90 dan 10	78%

**Pengujian Skenario ketiga**

Pada skenario pengujian pertama ini dilakukan untuk mengetahui tingkat akurasi yang dihasilkan menggunakan random state 60 menggunakan algoritma SVM. Setelah dilakukan pengujian di dapatkan hasil sebagai berikut.

**Tabel 4. Skenario Ketiga**

<b>Pembagian Data</b>	<b>Akurasi</b>
80 dan 20	78%
70 dan 30	70%
90 dan 10	54%

**KESIMPULAN DAN SARAN****KESIMPULAN**

Penelitian ini menunjukkan bahwa penggunaan algoritma Support Vector Machine (SVM) dengan kernel RBF (Radial Basis Function) dalam klasifikasi diabetes mampu mencapai hasil akurasi maksimal sebesar 87%. Hasil ini menunjukkan potensi besar algoritma SVM-RBF dalam mendeteksi diabetes secara efektif berdasarkan parameter kesehatan yang relevan.

SVM dengan kernel RBF memungkinkan model untuk menangkap hubungan yang kompleks dan nonlinier antara atribut kesehatan, sehingga meningkatkan kemampuan prediktifnya. Akurasi sebesar 87% mencerminkan tingkat keberhasilan model dalam mengklasifikasikan pasien diabetes dengan tepat.

**SARAN**

Meskipun hasil akurasi sebesar 87% sudah mencapai tingkat yang baik, ada beberapa saran yang dapat diberikan untuk meningkatkan kinerja model atau mendukung penelitian lebih lanjut:

1. Penggunaan Fitur Tambahan: Pertimbangkan penambahan fitur atau parameter kesehatan lainnya yang dapat memberikan kontribusi signifikan dalam mendeteksi

diabetes. Penggunaan fitur yang lebih kaya dapat meningkatkan kemampuan model untuk memahami kerumitan dataset.

2. Optimasi Parameter SVM-RBF: Lakukan optimasi parameter SVM-RBF seperti penentuan parameter C dan gamma secara cermat. Optimasi ini dapat membantu meningkatkan performa model dan mengurangi potensi overfitting atau underfitting.
3. Validasi dengan Dataset Eksternal: Lakukan validasi hasil dengan menggunakan dataset eksternal yang independen. Hal ini dapat memberikan kepercayaan tambahan terhadap generalisasi model ke populasi yang lebih luas.
4. Interpretasi Model: Lakukan analisis lebih lanjut terkait faktor-faktor yang paling berpengaruh dalam klasifikasi diabetes. Ini dapat membantu memahami faktor risiko dan kontribusi masing-masing parameter terhadap prediksi diabetes.
5. Kolaborasi dengan Profesional Kesehatan: Melibatkan profesional kesehatan dalam interpretasi hasil dan penggunaan model untuk mendukung keputusan klinis. Kolaborasi ini dapat meningkatkan relevansi dan penerapan model dalam praktik medis.

Dengan penerapan saran-saran di atas, diharapkan pengembangan model klasifikasi diabetes menggunakan SVM dengan kernel RBF dapat terus meningkatkan kinerjanya dan memberikan kontribusi positif dalam penanganan dini penyakit diabetes.

## **DAFTAR REFERENSI**

- [1]. Mahendra Dwifabri Purbolaksono. "Perbandingan Support Vector Machine dan Modified Balanced Random Forest dalam Deteksi Pasien Penyakit Diabetes". Available: <http://jurnal.iaii.or.id>.
- [2]. R.A. Wijayanti, M.T. Furqon, and S. Adinugroho. "Penerepan Algoritme Support Vector Machine Terhadap Klasifikasi Tingkat Risiko Pasien Gagal Ginjal". J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya, vol. 2, no. 10, pp. 3500–3507, 2018, [Online]. Available: <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/download/2647/991/>.
- [3]. E. Excavations, L. Classifiers, E. García-gonzalo, Z. Fernández-muñiz, P. José, and G. Nieto, "Hard-Rock Stability Analysis for Span Design in," pp. 1–19, 2016, doi: 10.3390/ma9070531.
- [4]. A. S. Nugroho, A. B. Witarto, and D. Handoko, "Support Vector Machine, Teori dan Aplikasinya dalam Bioinformatika," Proc. Indones. Sci. Meet. Cent. Japan, 2013, doi:10.1109/CCDC.2011.5968300.
- [5]. S. V. . Nugroho, "Paradigma Baru Dalam SoftComputing dan Aplikasinya," 2018.

- [6]. Hasan T. Abbas. 2019. "Predicting Long-term Type 2 Diabetes with *Support Vector Machine* using Oral Glucose Tolerance Test". **PLOS ONE**.
- [7]. Imelda A. Muis. 2015. "Penerapan Metode *Support Vector Machine* Menggunakan Kernel *Radial Basis Function* (RBF) pada Klasifikasi Tweet". **Jurnal Sains, Teknologi dan Industri** 12.
- [8]. Diniyal Amru Agatsa. 2020. "Klasifikasi Pasien Pengidap Diabetes menggunakan Metode *Support Vector Machine*". **Telkom University** 7.
- [9]. Purbolaksono. Mahendra Dwifabri. 2021. "Perbandingan *Support Vector Machine* dan *Modified Balanced Random Forest* dalam Deteksi Pasien Penyakit Diabetes". **Jurnal Rekayasa Sistem dan Teknologi Informasi** 5.
- [10]. Zahrah Chea. 2022. "Klasifikasi Menggunakan Metode *Support Vector Machine* dan *Random Forest* untuk Deteksi Awal Risiko Diabetes Melitus". **JURNAL GAUSSIAN UNDIP** 11.
- [11]. Andharini Dwi. 2019. "Klasifikasi Diabetes Mellitus menggunakan *Support Vector Machine* (Studi Kasus: Puskesmas Modopuro, Mojokerto)". **REKAYASA Universitas Trunojoyo** 12.
- [12]. Joshua Bonardo. 2021. "Perbandingan Akurasi Algoritma Decision Tree dan Algoritma *Support Vector Machine* pada Penyakit Diabetes". **Telkom University** 8.
- [13]. Aulia Suci. 2015. "Analisis Perbandingan KNN dengan SVM untuk Klasifikasi Penyakit Diabetes Retinopati berdasarkan Citra Eksudat dan Mikroaneurisma". **Jurnal ELKOMIKA Itenas** 3.
- [14]. Naisah Marito. 2022. "Komparasi Algoritma KNN dan Naïve Bayes untuk Klasifikasi Diagnosis Penyakit Diabetes Melitus". **Evolusi : Jurnal Sains dan Manajemen** 10.
- [15]. Januar Adi. 2016. "Klasifikasi Pengidap Diabetes pada Perempuan menggunakan Penggabungan Metode *Support Vector Machine* dan *K-Nearest Neighbour*". **Informatics Journal UNEJ** 1.
- [16]. Hairani. 2020. "K-means-SMOTE untuk menangani ketidakseimbangan kelas dalam klasifikasi penyakit diabetes dengan C4.5, SVM, dan naïve bayes". **Jurnal Teknologi dan Sistem Komputer UNDIP** 8.