

**ABSTRAK**  
**KLASIFIKASI PENYAKIT DIABETES MELITUS MENGGUNAKAN**  
**METODE *SUPPORT VECTOR MACHINE* (SVM)**

Rena Annisa'

Program Studi Teknik Informatika, Fakultas Teknik

Universitas Trunojoyo Madura

Pada tahun 2021, *International Diabetes Federation* (IDF) memprediksi bahwa 537 juta orang berusia antara 20 dan 79 tahun menderita diabetes di seluruh dunia. Penyakit diabetes juga menjadi penyebab 6,7 juta kematian di seluruh dunia dan diperkirakan akan terus bertambah setiap tahunnya. Oleh karena itu, untuk mengurangi angka penderita diabetes dibutuhkan suatu sistem yang menerapkan metode klasifikasi agar dapat mengklasifikasi penyakit diabetes. Metode yang digunakan dalam mengklasifikasi penyakit diabetes pada penelitian ini yaitu *Support Vector Machine* (SVM) menggunakan kernel *linear*, kernel *polynomial* dan kernel RBF. Adapun data yang digunakan bersumber dari *website kaggle* dengan jumlah data sebanyak 768 *record* dan 9 atribut termasuk kelas. Penelitian ini dilakukan guna mengetahui perbandingan hasil akurasi dari penerapan ketiga kernel tersebut pada metode SVM. Hasil akurasi tertinggi diperoleh dari penerapan kernel RBF. Akurasi yang diperoleh yaitu sebesar 85.42%. Sedangkan pada kernel *linear* dan kernel *polynomial* menghasilkan akurasi sebesar 80.56% dan 75%.

**Kata Kunci:** diabetes melitus, klasifikasi, *support vector machine*

**ABSTRACT**  
**CLASSIFICATION OF DIABETES MELLITUS USING SUPPORT VECTOR  
MACHINE (SVM) METHOD**

Rena Annisa'

*Department of Informatics Engineering, Faculty of Engineering*

*Trunojoyo University of Madura*

*In 2021, the International Diabetes Federation (IDF) predicts that 537 million people between the ages of 20 and 79 have diabetes worldwide. Diabetes is also the cause of 6.7 million deaths worldwide and is expected to continue to grow every year. Therefore, to reduce the number of diabetics, a system is needed that applies the classification method in order to classify diabetes. The method used in classifying diabetes in this study is the Support Vector Machine (SVM) using linear kernels, polynomial kernels and RBF kernels. The data used comes from the Kaggle website with a total of 768 records and 9 attributes including class. This research was conducted to compare the results of the accuracy of the application of the three kernels to the SVM method. The highest accuracy results are obtained from the application of RBF kernels. The accuracy obtained is equal to 85.42%. Meanwhile, the linear kernel and the polynomial kernel produced an accuracy of 80.56% and 75%.*

**Keywords:** *diabetes mellitus, classification, support vector machine*

## DAFTAR ISI

|                                                 |      |
|-------------------------------------------------|------|
| HALAMAN JUDUL.....                              | ii   |
| HALAMAN PENGESAHAN.....                         | iii  |
| HALAMAN PERNYATAAN ORISINALITAS.....            | iv   |
| KATA PENGANTAR .....                            | v    |
| DAFTAR ISI.....                                 | vii  |
| DAFTAR GAMBAR .....                             | x    |
| DAFTAR TABEL .....                              | xi   |
| ABSTRAK .....                                   | xiii |
| <i>ABSTRACT</i> .....                           | xiv  |
| BAB I PENDAHULUAN .....                         | 1    |
| 1.1    Latar Belakang .....                     | 1    |
| 1.2    Rumusan Masalah .....                    | 3    |
| 1.2.1    Permasalahan.....                      | 3    |
| 1.2.2    Metode Usulan .....                    | 3    |
| 1.2.3    Pertanyaan Penelitian .....            | 3    |
| 1.3    Batasan Masalah.....                     | 4    |
| 1.4    Tujuan Penelitian.....                   | 4    |
| 1.5    Manfaat Penelitian.....                  | 4    |
| 1.6    Sistematika Penulisan.....               | 4    |
| BAB II LANDASAN TEORI .....                     | 6    |
| 2.1    Penelitian Terdahulu.....                | 6    |
| 2.2    Landasan Teori .....                     | 12   |
| 2.2.1    Penyakit Diabetes.....                 | 12   |
| 2.2.2 <i>Data Mining</i> .....                  | 13   |
| 2.2.3 <i>Missing Value</i> .....                | 13   |
| 2.2.4 <i>Outlier</i> .....                      | 14   |
| 2.2.5    Normalisasi Data.....                  | 14   |
| 2.2.6 <i>Machine Learning</i> .....             | 15   |
| 2.2.7    Klasifikasi .....                      | 15   |
| 2.2.8 <i>Support Vector Machine (SVM)</i> ..... | 16   |

|                                   |                                                                     |    |
|-----------------------------------|---------------------------------------------------------------------|----|
| 2.2.9                             | Tahap Klasifikasi Menggunakan Metode SVM.....                       | 19 |
| 2.2.10                            | <i>Hyperparameter Tuning</i> Menggunakan <i>GridSearchCV</i> .....  | 20 |
| 2.2.11                            | Evaluasi Model.....                                                 | 21 |
| BAB III METODE USULAN.....        |                                                                     | 23 |
| 3.1                               | Kebutuhan Sistem.....                                               | 23 |
| 1.                                | <i>Hardware</i> .....                                               | 23 |
| 2.                                | <i>Software</i> .....                                               | 23 |
| 3.2                               | Alur Sistem.....                                                    | 23 |
| 3.3                               | Data Penelitian .....                                               | 25 |
| 3.4                               | <i>Pre-processing</i> Data .....                                    | 27 |
| 3.5                               | Perhitungan Manual Klasifikasi Menggunakan SVM.....                 | 28 |
| 3.5.1                             | <i>Toy Dataset</i> .....                                            | 28 |
| 3.5.2                             | <i>Pre-processing</i> Data.....                                     | 28 |
| 3.5.3                             | Pembagian Data .....                                                | 29 |
| 3.5.4                             | Pemodelan Klasifikasi SVM.....                                      | 29 |
| 3.6                               | Evaluasi Model.....                                                 | 38 |
| 3.7                               | Skenario Uji Coba .....                                             | 38 |
| BAB IV HASIL DAN PEMBAHASAN ..... |                                                                     | 40 |
| 4.1                               | Hasil Proses Pengelolaan Data Awal ( <i>Pre-processing</i> ) .....  | 40 |
| 4.1.1                             | Hasil Proses <i>Load Dataset</i> .....                              | 40 |
| 4.1.2                             | Hasil Proses Imputasi <i>Missing Value</i> .....                    | 41 |
| 4.1.3                             | Hasil Proses Deteksi <i>Outlier</i> .....                           | 43 |
| 4.1.4                             | Normalisasi Data.....                                               | 45 |
| 4.2                               | Hasil Proses Pembagian Data.....                                    | 46 |
| 4.3                               | Hasil Proses Tuning Parameter Menggunakan <i>GridSearchCV</i> ..... | 47 |
| 4.3.1                             | Hasil Proses Tuning Parameter Pada Kernel <i>Linear</i> .....       | 47 |
| 4.3.2                             | Hasil Proses Tuning Parameter Pada Kernel <i>Polynomial</i> .....   | 48 |
| 4.3.3                             | Hasil Proses Tuning Parameter Pada Kernel RBF .....                 | 49 |
| 4.4                               | Hasil Skenario Uji Coba.....                                        | 51 |
| 4.4.1                             | Hasil Skenario Uji Coba 1.....                                      | 51 |
| 4.4.2                             | Hasil Skenario Uji Coba 2.....                                      | 53 |
| 4.4.3                             | Hasil Skenario Uji Coba 3.....                                      | 56 |

|                        |                                       |    |
|------------------------|---------------------------------------|----|
| 4.5                    | Hasil Analisis Skenario Uji Coba..... | 59 |
| BAB V PENUTUP.....     |                                       | 60 |
| 5.1                    | Kesimpulan.....                       | 60 |
| 5.2                    | Saran.....                            | 60 |
| REFERENSI .....        |                                       | 61 |
| BIOGRAFI PENULIS ..... |                                       | 65 |

## DAFTAR GAMBAR

|                                                                                                |    |
|------------------------------------------------------------------------------------------------|----|
| <b>Gambar 2. 1</b> <i>Hyperplane</i> pada SVM .....                                            | 16 |
| <b>Gambar 2. 2</b> Tahap klasifikasi metode SVM .....                                          | 19 |
| <b>Gambar 3. 1</b> Alur sistem.....                                                            | 24 |
| <b>Gambar 3. 2</b> <i>Screenshot</i> dataset diabetes.....                                     | 26 |
| <b>Gambar 4. 1</b> Contoh 5 data teratas diabetes dataset .....                                | 40 |
| <b>Gambar 4. 2</b> Hasil cek <i>missing value</i> sebelum proses imputasi .....                | 41 |
| <b>Gambar 4. 3</b> Hasil proses mengganti nilai 0 menjadi NaN .....                            | 41 |
| <b>Gambar 4. 4</b> Hasil cek <i>missing value</i> setelah nilai 0 diganti NaN.....             | 42 |
| <b>Gambar 4. 5</b> Hasil rata-rata atribut pada setiap kelas .....                             | 42 |
| <b>Gambar 4. 6</b> Program dan <i>output</i> setelah dilakukan imputasi.....                   | 43 |
| <b>Gambar 4. 7</b> Hasil nilai <i>z-score</i> .....                                            | 43 |
| <b>Gambar 4. 8</b> Hasil data yang terdeteksi <i>outlier</i> .....                             | 44 |
| <b>Gambar 4. 9</b> Jumlah data setelah dibersihkan dari <i>outlier</i> .....                   | 44 |
| <b>Gambar 4. 10</b> Beberapa atribut yang terdeteksi <i>oulier</i> dalam satu baris data ..... | 44 |
| <b>Gambar 4. 11</b> Jumlah data pada setiap kelas setelah dibersihkan dari <i>outlier</i> .... | 45 |
| <b>Gambar 4. 12</b> Hasil normalisasi data.....                                                | 45 |
| <b>Gambar 4. 13</b> <i>Output</i> program tuning parameter pada kernel <i>linear</i> .....     | 48 |
| <b>Gambar 4. 14</b> <i>Output</i> program tuning parameter pada kernel <i>polynomial</i> ..... | 49 |
| <b>Gambar 4. 15</b> <i>Output</i> program tuning parameter pada kernel RBF .....               | 50 |
| <b>Gambar 4. 16</b> Hasil <i>confusion matrix</i> pada kernel <i>linear</i> .....              | 52 |
| <b>Gambar 4. 17</b> Hasil evaluasi model pada kernel <i>linear</i> .....                       | 53 |
| <b>Gambar 4. 18</b> Visualisasi <i>hyperplane</i> pada kernel <i>linear</i> .....              | 53 |
| <b>Gambar 4. 19</b> Hasil <i>confusion matrix</i> pada kernel <i>polynomial</i> .....          | 55 |
| <b>Gambar 4. 20</b> Hasil evaluasi model pada kernel <i>polynomial</i> .....                   | 55 |
| <b>Gambar 4. 21</b> Visualisasi <i>hyperplane</i> pada kernel <i>polynomial</i> .....          | 56 |
| <b>Gambar 4. 22</b> Hasil <i>confusion matrix</i> pada kernel RBF.....                         | 57 |
| <b>Gambar 4. 23</b> Hasil evaluasi model pada kernel RBF.....                                  | 58 |
| <b>Gambar 4. 24</b> Visualisasi <i>hyperplane</i> pada kernel RBF.....                         | 58 |

## DAFTAR TABEL

|                                                                                                            |    |
|------------------------------------------------------------------------------------------------------------|----|
| <b>Tabel 2. 1</b> Rangkuman jurnal penelitian terdahulu .....                                              | 9  |
| <b>Tabel 2. 2</b> Keterangan parameter.....                                                                | 21 |
| <b>Tabel 2. 3</b> <i>Confusion matrix</i> .....                                                            | 22 |
| <b>Tabel 3. 1</b> Atribut dataset penyakit diabetes .....                                                  | 27 |
| <b>Tabel 3. 2</b> Nilai normal atribut .....                                                               | 27 |
| <b>Tabel 3. 3</b> <i>Toy</i> dataset diabetes.....                                                         | 28 |
| <b>Tabel 3. 4</b> Hasil perhitungan <i>z-Score</i> .....                                                   | 29 |
| <b>Tabel 3. 5</b> Data <i>training</i> .....                                                               | 29 |
| <b>Tabel 3. 6</b> Data <i>testing</i> .....                                                                | 29 |
| <b>Tabel 3. 7</b> <i>Transpose</i> data.....                                                               | 29 |
| <b>Tabel 3. 8</b> Perhitungan <i>dot product</i> kernel <i>linear</i> .....                                | 30 |
| <b>Tabel 3. 9</b> Matriks hessian.....                                                                     | 30 |
| <b>Tabel 3. 10</b> Hasil perhitungan nilai error.....                                                      | 31 |
| <b>Tabel 3. 11</b> Hasil perhitungan delta alpha .....                                                     | 31 |
| <b>Tabel 3. 12</b> Hasil perhitungan alpha baru.....                                                       | 31 |
| <b>Tabel 3. 13</b> Hasil perhitungan nilai bias .....                                                      | 31 |
| <b>Tabel 3. 14</b> Hasil perhitungan <i>dot product</i> data <i>training</i> dan data <i>testing</i> ..... | 32 |
| <b>Tabel 3. 15</b> <i>Transpose</i> data.....                                                              | 32 |
| <b>Tabel 3. 16</b> Perhitungan <i>dot product</i> kernel <i>polynomial</i> .....                           | 33 |
| <b>Tabel 3. 17</b> Matriks hessian.....                                                                    | 33 |
| <b>Tabel 3. 18</b> Hasil perhitungan nilai error.....                                                      | 34 |
| <b>Tabel 3. 19</b> Hasil perhitungan delta alpha .....                                                     | 34 |
| <b>Tabel 3. 20</b> Hasil perhitungan alpha baru.....                                                       | 34 |
| <b>Tabel 3. 21</b> Hasil perhitungan nilai bias .....                                                      | 34 |
| <b>Tabel 3. 22</b> Hasil perhitungan <i>dot product</i> data <i>training</i> dan data <i>testing</i> ..... | 35 |
| <b>Tabel 3. 23</b> Perhitungan <i>dot product</i> kernel RBF .....                                         | 36 |
| <b>Tabel 3. 24</b> Matriks hessian.....                                                                    | 36 |
| <b>Tabel 3. 25</b> Hasil perhitungan nilai error.....                                                      | 36 |
| <b>Tabel 3. 26</b> Hasil perhitungan delta alpha .....                                                     | 37 |
| <b>Tabel 3. 27</b> Hasil perhitungan alpha baru.....                                                       | 37 |
| <b>Tabel 3. 28</b> Hasil perhitungan nilai bias .....                                                      | 37 |

|                                                                                                            |    |
|------------------------------------------------------------------------------------------------------------|----|
| <b>Tabel 3. 29</b> Hasil perhitungan <i>dot product</i> data <i>training</i> dan data <i>testing</i> ..... | 37 |
| <b>Tabel 3. 30</b> Skenario uji coba .....                                                                 | 38 |
| <b>Tabel 3. 31</b> Nilai parameter dan iterasi maksimal .....                                              | 39 |
| <b>Tabel 4. 1</b> Program <i>load dataset</i> .....                                                        | 40 |
| <b>Tabel 4. 2</b> Program pembagian <i>dataset</i> .....                                                   | 46 |
| <b>Tabel 4. 3</b> Hasil pembagian <i>dataset</i> .....                                                     | 46 |
| <b>Tabel 4. 4</b> Hasil distribusi data .....                                                              | 46 |
| <b>Tabel 4. 5</b> Program tuning parameter pada kernel <i>linear</i> .....                                 | 47 |
| <b>Tabel 4. 6</b> Hasil tuning pada kernel <i>linear</i> .....                                             | 48 |
| <b>Tabel 4. 7</b> Program tuning parameter pada kernel <i>polynomial</i> .....                             | 48 |
| <b>Tabel 4. 8</b> Hasil tuning pada kernel <i>polynomial</i> .....                                         | 49 |
| <b>Tabel 4. 9</b> Program tuning parameter pada kernel RBF .....                                           | 50 |
| <b>Tabel 4. 10</b> Hasil tuning pada kernel RBF .....                                                      | 50 |
| <b>Tabel 4. 11</b> Program evaluasi pemodelan pada kernel <i>linear</i> .....                              | 51 |
| <b>Tabel 4. 12</b> Program evaluasi pemodelan pada kernel <i>polynomial</i> .....                          | 54 |
| <b>Tabel 4. 13</b> Program evaluasi pemodelan pada kernel RBF .....                                        | 56 |
| <b>Tabel 4. 14</b> Hasil perbandingan akurasi tiga fungsi kernel .....                                     | 59 |