**PREDIKSI PENYAKIT DIABETES MENGGUNAKAN METODE ALGORITMA SUPPORT VECTOR MACHINE (SVM)**

**SKRIPSI**

Diajukan untuk Memenuhi Salah Satu Syarat Kelulusan

Jenjang Strata 1 pada Program Studi Teknik Informatika

Univesitas Jenderal Achmad Yani

Disusun Oleh

**Hovi Sohibul Wafa**

**NIM. 3411171171**



**PROGRAM STUDI INFORMATIKA**

**FAKULTAS SAINS DAN INFORMATIKA**

**UNIVERSITAS JENDERAL ACHMAD YANI**

**CIMAHI**

**2022**

# LEMBAR PENGESAHAN

**PREDIKSI PENYAKIT DIABETES MENGGUNAKAN**

**METODE ALGORITMA SUPPORT VECTOR MACHINE**

**(SVM)**

**“Setelah membaca skripsi ini dengan seksama, menurut pertimbangan kami telah memenuhi persyaratan ilmiah sebagai suatu skripsi”**

**Cimahi,**

**Hovi Sohibul Wafa**

**NIM. 3411 1711 71**

|  |  |
| --- | --- |
| Pembimbing 1  **Asep Id Hadiana, S.SI., M.KOM.**  **NID. 4221 800 78** | Pembimbing 2  **Fajri Rakhmat U, S.T., M.T.**  **NID. 4121 858 88** |

**Mengetahui,**

|  |  |
| --- | --- |
| **Dekan Fakultas**  **Sains dan Informatika** | **Ketua Jurusan**  **Informatika** |
| **Dr. Anceu Murniati, S.Si., M.Si.** | **Agus Komarudin S.Kom., M.T** |
| **NID. 4121 263 69** | **NID. 4121 758 78** |

# PERNYATAAN KEASLIAN

Saya menyatakan dengan ini bahwa skripsi dengan judul **“PREDIKSI PENYAKIT DIABETES MENGGUNAKAN METODE ALGORITMA SUPPORT VECTOR MACHINE (SVM)”** ini beserta isinya adalah benar-benar karya saya dan tidak melakukan penjiplakan atau pengutipan dengan cara yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan.

Atas pernyataan ini saya siap menanggung resiko atau sanksi yang dijatuhkan kepada saya apabila kemudian ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya atau ada klaim dari pihak lain terhadap keaslian karya saya ini. Demikian pernyataan saya buat,

|  |
| --- |
| Cimahi  Yang Membuat Pernyataan    **Hovi Sohibul Wafa**  **NIM. 3411 1711 71** |

# ABSTRAK

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 masyarakat diberbagai negara terkena penyakit diabetes termasuk di Indonesia. Penyebab utama diabetes masih belum diketahui, namun banyak yang percaya bahwa faktor *genetic* 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 gab tersebut bahwa dengan menggunakan metode algoritma SVM *Radial Basis Function* (RBF) berbasis *Forward Selection* 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*.

**Kata Kunci:** *Diabetes Mellitus*, *Prediksi*, *Support Vector Machine, KFold, Confusion Matrix, Mean Square Error*

# ABSTRACT

*Diabetes Mellitus (DM) or better known as diabetes is a chronic disease caused by the failure of the pancreas to produce adequate amounts of the hormone insulin, causing an increase in blood glucose levels. Diabetes Mellitus is a dangerous disease, many countries are affected by diabetes, including in Indonesia. The main cause of diabetes is still unknown, but many believe that genetics and lifestyle factors can play a major role in diabetes. Researchers in the field of bioinformatics have been trying to tackle this disease and create systems to assist in diabetes prediction. From various existing studies, many use methods such as C4.5, KNN, Naïve Bayes, and Linear SVM in building systems, but the SVM Radial Basis Function (RBF) method is rarely used because the accuracy results obtained are not sufficient to be used in diabetes prediction systems. . In this study, the answer is that using the SVM Radial Basis Function (RBF) algorithm method can produce high accuracy by reaching 91%. Accuracy testing was carried out using the Confusion Matrix and Mean Square Error forecasting with kfold multiples of 10. This study aims to determine whether the patient/patient can get diabetes or not by applying data mining and classification techniques using the SVM Radial Basis Function algorithm based on Forward Selection.*

***Keywords***: *Diabetes Mellitus*, *Prediction*, *Support Vector Machine, KFold, Confusion Matrix, Mean Square Error*

# KATA PENGANTAR

Segala puji dan syukur kepada Allah SWT akan karunia-Nya yang berlimpah dan telah memberikan kemudahan serta kelancaran, alhamdullilah penulis dapat menyelesaikan penulisan skripsi dengan judul: **“PREDIKSI PENYAKIT DIABETES MENGGUNAKAN METODE ALGORITMA SUPPORT VECTOR MACHINE (SVM)”**.

Penulis menyadari bahwa skripsi ini masih banyak kekurangan dan jauh dari kata sempurna karena menyadari keterbatasannya. Oleh sebab itu penulis sangat membutuhkan saran dan kritikan dengan sifat membangun

|  |
| --- |
| Cimahi  Yang Membuat Pernyataan    **Hovi Sohibul Wafa**  **NIM. 3411 1711 71** |

# UCAPAN TERIMA KASIH

Selama proses penyusunan skripsi ini, penulis mendapatkan banyak hambatan dan rintangan namun karena berkat karunia-Nya penulis dapat menyelesaikan skripsi ini. Oleh karena itu dengan penuh ketulusan dan kerendahan hati. Penulis mengucapkan terima kasih banyak kepada:

1. ALLAH SWT, Yang telah memberikan berkat dan rahmat-Nya serta kelancaran.
2. Juga kedua orang tua, Mamah Serta Ayah, yang telah memberikan dukungan dan semangat dalam menyelesaikan tugas akhir ini serta selalu mendoakan saya agar tetap sehat secara jasmani maupun rohani.
3. Kakak, yang selalu memberikan dukungan disaat sedang ada hambatan dan selalu mengingatkan saya untuk tidak menyerah dan selalu berusaha.
4. Yth. Ibu Dr. Anceu Murniati, S.SI., M. SI. selaku Dekan Fakultas Sains dan Informatika Universitas Jenderal Achmad Yani.
5. Yth. Ibu Wina Witanti, S.T., M.T selaku Ketua Jurusan Teknik Informatika Universitas Jenderal Achmad Yani.
6. Yth. Bapak Asep Id Hadiana S.SI., M.KOM, selaku dosen pembimbing I yang dengan sabar membimbing, memberikan ide, solusi, dan arahan, serta memberikan motivasi dan bisa meluangkan waktunya disaat jam sibuk.
7. Yth. Bapak Fajri Rakhmat Umbara S.T., M.T, selaku dosen pembimbing II yang dengan sabar membimbing, memberikan ide, solusi, dan arahan, serta memberikan motivasi dan bisa meluangkan waktunya disaat jam sibuk.
8. Kepada semua dosen beserta staff karyawan di Jurusan Teknik Informatika Universitas Jenderal Achmad Yani.
9. Kepada keluarga besar Fakultas Sains dan Informatika Universitas Jenderal Achmad Yani, dari mulai dosen, karyawan, hingga staff tu Fakultas Sains dan Informatika.
10. Juga kepada teman-teman yang sudah memberikan semangat, dan berjuang hingga sampai sekarang.

# DAFTAR ISI

[LEMBAR PENGESAHAN i](#_Toc91265895)

[PERNYATAAN KEASLIAN ii](#_Toc91265896)

[ABSTRAK iii](#_Toc91265897)

[ABSTRACT iv](#_Toc91265898)

[KATA PENGANTAR v](#_Toc91265899)

[UCAPAN TERIMA KASIH vi](#_Toc91265900)

[DAFTAR ISI vii](#_Toc91265901)

[DAFTAR GAMBAR x](#_Toc91265902)

[DAFTAR TABEL xii](#_Toc91265903)

[DAFTAR SIMBOL xiii](#_Toc91265904)

[BAB I 1](#_Toc91265905)

[PENDAHULUAN 1](#_Toc91265906)

[1.1 Latar Belakang 1](#_Toc91265907)

[1.2 Rumusan Masalah 4](#_Toc91265908)

[1.3 Ruang Lingkup 4](#_Toc91265909)

[1.4 Tujuan Penelitian 4](#_Toc91265910)

[1.5 Luaran dan Manfaat 4](#_Toc91265911)

[1.6 Metode Penelitian 4](#_Toc91265912)

[1.6.1. Data Collection 5](#_Toc91265913)

[1.6.2. Pre-Processing 5](#_Toc91265914)

[1.6.3. Proses Prediksi 6](#_Toc91265915)

[1.6.4. Pembangunan Perangkat Lunak 6](#_Toc91265916)

[1.6.5. Pengujian & Evaluasi 7](#_Toc91265917)

[1.6.6. Pelaporan & Publikasi 7](#_Toc91265918)

[1.7 Sistematik Penulisan 7](#_Toc91265919)

[BAB II 9](#_Toc91265920)

[TINJAUAN PUSTAKA 9](#_Toc91265921)

[2.1. Data Mining 10](#_Toc91265922)

[2.2. Diabetes Melitus 11](#_Toc91265923)

[2.3. Support Vector Machine (Radial Basis Function) 12](#_Toc91265924)

[2.4. Forward Selection 14](#_Toc91265925)

[2.5. Confusion Matrix 15](#_Toc91265926)

[2.6. Mean Square Error 16](#_Toc91265927)

[BAB III 18](#_Toc91265928)

[ANALISIS DAN PERANCANGAN 18](#_Toc91265929)

[3.1. Analisis Prediksi Diabetes Menggunakan Support Vector Machine 18](#_Toc91265930)

[3.2. Data Collection 20](#_Toc91265931)

[3.2.1. Deskripsi Atribut Data 20](#_Toc91265932)

[3.2.2. Data Pasien Wanita Keturunan Indian Pima 22](#_Toc91265933)

[3.3. Pre Processing 22](#_Toc91265934)

[3.5.9.1. Data Cleaning & Selection 22](#_Toc91265935)

[a. Data Tranformasi 23](#_Toc91265936)

[3.4. Proses Prediksi Diabetes 24](#_Toc91265937)

[3.5. Perancangan Perangkat Lunak 26](#_Toc91265938)

[3.5.1. Analisis Kebutuhan Perangkat Lunak 26](#_Toc91265939)

[3.5.2. Deskripsi Aktor Sistem Prediksi 26](#_Toc91265940)

[3.5.3. Use Case Diagram 26](#_Toc91265941)

[3.5.4. Deskripsi Use Case 27](#_Toc91265942)

[3.5.5. Skenario Use Case 28](#_Toc91265943)

[3.5.6. Activity Diagram 35](#_Toc91265944)

[3.5.7. Sequence Diagram 40](#_Toc91265945)

[3.5.8. Class Diagram 44](#_Toc91265946)

[3.5.9. Perancangan Database 45](#_Toc91265947)

[3.5.10. Perancangan Antarmuka 46](#_Toc91265948)

[BAB IV 52](#_Toc91265949)

[IMPLEMENTASI DAN PENGUJIAN 52](#_Toc91265950)

[4.1. Implementasi Sistem 52](#_Toc91265951)

[4.2. Implementasi Database 52](#_Toc91265952)

[4.3. Implementasi Antarmuka 53](#_Toc91265953)

[4.4. Pengujian Akurasi 57](#_Toc91265954)

[BAB V 59](#_Toc91265955)

[KESIMPULAN DAN SARAN 59](#_Toc91265956)

[DAFTAR ISI 60](#_Toc91265957)

# DAFTAR GAMBAR

[**Gambar 1. 1 Metode Penelitian Prediksi Diabetes** 5](#_Toc90293451)

[**Gambar 2. 1 Tahapan-tahapan proses dalam data mining** 11](#_Toc90293463)

[**Gambar 2. 2 Hyperplane** 13](#_Toc90293464)

[**Gambar 3. 1 Perancangan Sistem Prediksi Diabetes** 20](#_Toc90293479)

[**Gambar 3. 2 Use Case Prediksi Diabetes** 27](#_Toc90293480)

[**Gambar 3. 3 Activity Lihat Processing Data** 36](#_Toc90293481)

[**Gambar 3. 4 Activity Lihat Data Latih** 37](#_Toc90293482)

[**Gambar 3. 5 Activity Lihat Data Uji** 38](#_Toc90293483)

[**Gambar 3. 6 Activity Lihat Hasil Evaluasi** 39](#_Toc90293484)

[**Gambar 3. 7 Activity Tambah Data** 40](#_Toc90293485)

[**Gambar 3. 8 Sequence Processing Data** 41](#_Toc90293486)

[**Gambar 3. 9 Sequence Data Latih** 42](#_Toc90293487)

[**Gambar 3. 10 Sequence Data Uji** 42](#_Toc90293488)

[**Gambar 3. 11 Sequence Evaluasi** 43](#_Toc90293489)

[**Gambar 3. 12 Sequence Tambah Data** 44](#_Toc90293490)

[**Gambar 3. 13 Class Prediksi Diabetes** 44](#_Toc90293491)

[**Gambar 3. 14 Antarmuka Home** 46](#_Toc90293492)

[**Gambar 3. 15 Antarmuka Processing Data** 47](#_Toc90293493)

[**Gambar 3. 16 Antarmuka Data Latih** 48](#_Toc90293494)

[**Gambar 3. 17 Antarmuka Data Uji** 49](#_Toc90293495)

[**Gambar 3. 18 Antarmuka Evaluasi Data** 50](#_Toc90293496)

[**Gambar 3. 19 Antarmuka Tambah Data** 51](#_Toc90293497)

[**Gambar 4. 1 Database Pasien Diabetes** 53](#_Toc90293508)

[**Gambar 4. 2 Implementasi Antarmuka Home** 53](#_Toc90293509)

[**Gambar 4. 3 Implementasi Antarmuka Processing Data** 54](#_Toc90293510)

[**Gambar 4. 4 Implementasi Antarmuka Data Latih** 55](#_Toc90293511)

[**Gambar 4. 5 Implementasi Antarmuka Data Uji** 56](#_Toc90293512)

[**Gambar 4. 6 Implementasi Antarmuka Evaluasi Data** 57](#_Toc90293513)

[**Gambar 4. 7 Implementasi Antarmuka Tambah Data** 57](#_Toc90293514)

# DAFTAR TABEL

[**Tabel 3. 1 Deskripsi Atribut Data** 20](#_Toc90293413)

[**Tabel 3. 2 Contoh Data Pemeriksaan Pasien Wanita Keturunan Indian Pima** 1](#_Toc90293414)

[**Tabel 3. 3 Data Pasien Wanita Indian Pima** 21](#_Toc90293415)

[**Tabel 3. 4 Contoh Transformation Dataset** 1](#_Toc90293416)

[**Tabel 3. 5 Deskrips Aktor** 26](#_Toc90293417)

[**Tabel 3. 6 Deskripsi Use Case** 27](#_Toc90293418)

[**Tabel 3. 7 Skenario Lihat Processing Data** 28](#_Toc90293419)

[**Tabel 3. 8 Skenario Lihat Data Latih** 30](#_Toc90293420)

[**Tabel 3. 9 Skenario Lihat Data Uji** 31](#_Toc90293421)

[**Tabel 3. 10 Skenario Lihat Hasil Evaluasi** 32](#_Toc90293422)

[**Tabel 3. 11 Skenario Tambah Data** 34](#_Toc90293423)

[**Tabel 3. 12 Tabel Pasien** 45](#_Toc90293424)

[**Tabel 4. 1 Tabel Nilai Confusion Matrix** 58](#_Toc90293425)

# DAFTAR SIMBOL

1. **Simbol Use Case Diagram**

**Tabel 1 Tabel Use Case Diagram**

|  |  |  |  |
| --- | --- | --- | --- |
| **No.** | **Simbol** | **Nama Simbol** | **Kerangan** |
| 1. |  | *Actor* | Menspesifikasikan himpuan peran yang pengguna mainkan ketika berinteraksi dengan use case. |
| 2. |  | *Use Case* | Deskripsi dari urutan aksi-aksi yang ditampilkan sistem yang menghasilkan suatu hasil yang terukur bagi suatu aktor. |
| 3. |  | *Association* | Apa yang menghubungkan antara objek satu dengan objek lainnya. |
| 4. |  | *Extend* | Menspesifikasikan bahwa use casetarget memperluas perilaku dari use case sumber pada suatu titik yang diberikan. |
| 5. |  | *Include* | Menspesifikasikan bahwa use case sumber secara eksplisit. |

(Sumber: Munawar, Pemodelan Visual UML, 2005)

1. **Simbol Activity Diagram**

**Tabel 2 Tabel Activity Diagram**

|  |  |  |  |
| --- | --- | --- | --- |
| **No.** | **Simbol** | **Nama Simbol** | **Keterangan** |
| 1. |  | *Activity* | Memperlihatkan bagaimana masing-masing kelas antarmuka saling berinteraksi satu sama lain. |
| 2. |  | *Action* | State dari sistem yang mencerminkan eksekusi dari suatu aksi. |
| 3. |  | *Initial Node* | Bagaimana objek dibentuk atau diawali. |
| 4. |  | *Activity Final Node* | Bagaimana objek dibentuk dan diakhiri. |
| 5. |  | *Decision* | Digunakan untuk menggambarkan suatu keputusan/tindakan yang harus diambil pada kondisi tertentu. |
| 6. |  | *Line Connector* | Digunakan untuk menghubungkan satu simbol dengan simbol yang lainnya. |

(Sumber: Munawar, Pemodelan Visual UML, 2005)

1. **Simbol Sequence Diagram**

**Tabel 3 Tabel Sequence Diagram**

|  |  |  |  |
| --- | --- | --- | --- |
| **No.** | **Simbol** | **Nama Simbol** | **Keterangan** |
| 1. |  | *LifeLine* | Objek entity, antarmuka yang saling berinteraksi. |
| 2 |  | *Message* | Spesifikasi dari komunikasi antar objek yang memuat informasi-informasi tentang aktivitas yang terjadi. |
| 3. |  | *Message* | Spesifikasi dari komunikasi antar objek yang memuat informasi-informasi tentang aktivitas yang terjadi. |

(Sumber: Munawar, Pemodelan Visual UML, 2005)

1. **Simbol Class Diagram**

**Tabel 4 Tabel Class Diagram**

|  |  |  |  |
| --- | --- | --- | --- |
| **No.** | **Gambar** | **Nama** | **Keterangan** |
| 1 |  | Generalization | Hubungan dimana objek (descendent) berbagi perilaku dan struktur data dari objek yang ada di atasnya atau objek induk (ancestor). |
| 2 |  | Class | Himpunan dari objek-objek yang berbagi atribut serta operasi yang sama. |
| 3 |  | Association | Apa yang menghubungkan antara objek satu dengan objek lainnya. |
| 4 |  | Dependency | Hubungan dimana perubahan yang terjadi pada suatu elemen mandiri (independent) akan mempegaruhi elemen yang bergantung padanya elemen yang tidak mandiri. |

(Sumber : Munawar, Pemodelan Visual UML, 2005)

# BAB I

# PENDAHULUAN

## Latar Belakang

Diabetes Melitus 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. Penyakit Diabetes Melitus dikenal sebagai penyakit *silent killer* dikarenakan para penderita tidak menyadari dan saat diketahui sudah terjadi komplikasi. Penyakit ini dapat menyerang hampir seluruh sistem tubuh manusia, mulai dari kulit sampai jantung yang menimbulkan komplikasi [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].

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 [3].

Penyebab utama diabetes masih belum diketahui, namun para dokter percaya bahwa pada kondisi ini diduga berkaitan dengan faktor *genetic* (keturunan) dan gaya hidup memainkan peran utama dalam penyakit tersebut. Penderita diabetes menghadapi resiko terkena beberapa masalah kesehatan sekunder seperti penyakit jantung dan kerusakan saraf [4]. Untuk mengendalikan peningkatan terkena diabetes mellitus perlunya mendiagnosis penyakit tersebut secara dini untuk mencegah komplikasi dan mengurangi resiko masalah kesehatan yang terbilang parah [5] [6]. Namun, para dokter harus menganalisa banyak faktor sebelum mendiagnosis yang membuat pekerjaan dokter menjadi sulit. Ada cara yang dapat dilakukan untuk membantu pekerjaan dokter, yaitu membuat suatu prediksi/mendeteksi penyakit diabetes. Teknik yang diperlukan dalam prediksi yaitu dengan *data mining*. *Data mining* sebagai proses seleksi, eksplorasi, dan pemodelan dari sejumlah data besar untuk menghasilkan sebuah pengetahuan [7]. 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 [8]. Alhasil dengan teknik tersebut menghasilkan sebuah sistem yang dapat memprediksi penyakit tersebut. Banyaknya penelitian yang menggunakan solusi ini dengan berbagai metode algoritma. Dari menggunakan deep learning yaitu Neural Network [9] hingga menggunakan machine learning seperti C4.5, *Naive Bayes*, *Logistic Regression* dan SVM *Linier* [10] [11] [12]. Akan tetapi dari semua metode algoritma yang digunakan, belum ada penelitian menggunakan metode SVM *Radial Basis Function* berbasis *Forward Selection* dalam prediksi diabetes. Dan pada penelitian ini menjawab gab tersebut bahwa dengan menggunakan metode algoritma SVM *Radial Basis Function* (RBF) berbasis *Forward Selection* dapat menghasilkan akurasi yang tinggi.

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 [13]. Oleh karena itu, dibutuhkan kernel untuk mentransformasikan data ke ruang dimensi yang lebih tinggi yang disebut ruang kernel. *Radial Basis Function* (RBF)adalah salah satu kernel yang 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 [13] [14].

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 [15]. Dari penjelasan singkat disimpulkan bahwa metode SVM *Radial Basis Function* (RBF) dan *Forward Selection* sangat cocok untuk prediksi penyakit diabetes. Hasil metode yang digunakan penelitian ini dievaluasi dengan *k-fold cross validation* dengan nilai cv 10 serta dilakukan pengujian dengan teknik perhitungan *Confusion Matrix* dan *Mean Square Error* [16][17].

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*. Harapan pada penelitian ini agar dataset yang telah dibagi menjadi data latih dan data uji dapat terprediksi dengan baik menggunakan SVM *Radial Basis Function* berbasis *Forward Selection* serta akurasi yang dihasilkan dengan pengujian *Confusion Matrix* dan MSE menghasilkan nilai yang tinggi.

## Rumusan Masalah

Berdasarkan masalah yang dijelaskan, diabetes melitus adalah penyakit yang tidak disadari dan saat sudah diketahui sudah berkomplikasi. Oleh sebab itu, perlunya mendeteksi penyakit tersebut secara dini dengan membuat sebuah model yang mampu memprediksi penyakit tersebut berdasarkan hasil pemeriksaan seperti glukosa, insulin, berat badan, ketebalan kulit, dll.

## Ruang Lingkup

Data yang digunakan dalam penelitian ini, adalah dataset pasien wanita keturunan indian pima yang didapatkan pada website UCI Machine Learning Repository.

## Tujuan Penelitian

Merancang dan membangun mesin yang mampu memprediksi penyakit diabetes dari hasil pemeriksaan kesehatan dengan teknik data mining dan perhitungan secara klasifikasi dengan bantuan algoritma SVM *Radial Basis Function* dan metode *Forward Selection* .

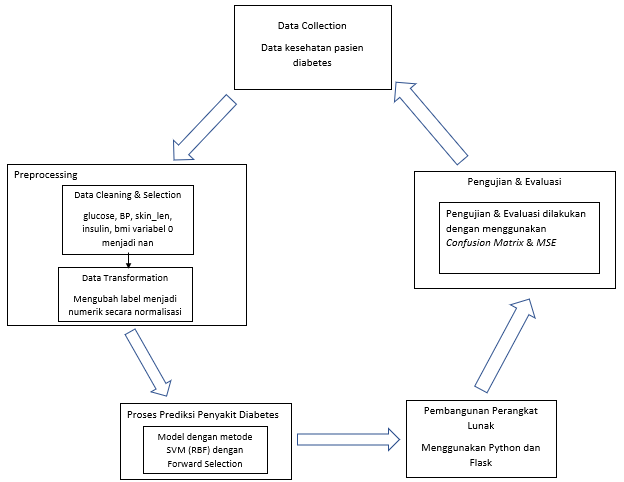
## Luaran dan Manfaat

Luaran dari penelitian ini adalah memprediksi penyakit diabetes menggunakan data hasil pemeriksaan pasien menggunakan teknik klasifikasi dengan metode SVM *Radial Basis Function* berbasis *Forward Selection*.

Manfaat yang diperoleh adalah dengan metode SVM *Radial Basis Function* berbasis *Forward Selection* mampu memprediksi penyakit diabetes dari hasil pemeriksaan kesehatan. Dengan menghadirkan sistem prediksi dapat membantu menekan jumlah penderita diabetes.

## Metode Penelitian

Metode penelitian ini berisi tahapan-tahapan yang dilakukan dalam membangun sistem prediksi diabetes, dimulai dari mengumpulkan data, selanjutnya melakukan tahapan *preprocessing* dalam pengolahan data, kemudian membuat model prediksi dengan metode SVM *Radial Basis Function* yang diusulkan, selanjutnya melakukan pengujian yaitu menghitung akurasi yang didapatkan pada model prediksi SVM *Radial Basis Function* dengan menggunakan *Confusion Matrix* dan *Mean Square Error* yang dapat dilihat pada Gambar 1.1 Metode Penelitian Prediksi Diabetes.



**Gambar 1. 1 Metode Penelitian Prediksi Diabetes**

### Data Collection

Pada tahapan ini, melakukan pengumpulan data yang diambil pada website kaggle. Data yang digunakan adalah data wanita dengan keturunan india pima.

### Pre-Processing

Pada tahapan ini adalah mengubah data pasien yang sudah tersedia untuk di proses menjadi data yang siap untuk digunakan. Pemrosesan data yang dilakukan pada penelitian ini meliputi: *Data Transformasi*, *Selection*.

1. **Data Cleaning**

Dalam tahapan ini, data pasien yang telah tersedia melakukan cleaning data. Dengan cara mengubah nilai 0 pada atribut glukosa, bp, skin len, insulin, bmi menjadi nan agar dapat diatasi.

1. **Data Selection**

Setelah atribut data yang dipilih telah diubah, lalu memilih data dengan atribut pregnancies, glukosa, bp, skin len, insulin, bmi, pedigree fun serta age dengan melakukan filtering pada atribut yang dipilih berdasarkan total yang telah dihitung dengan rumus median.

### Proses Prediksi

Pada tahapan ini, data yang telah diproses pada tahapan *pre-processing* kemudian melakukan drop pada atribut output. Setelah melakukan drop atribut output tersebut, kemudian melakukan pembagian data menjadi data uji dan data latih. Lalu tiap nilai yang ada pada atribut pregnancies, glucose, bp, skin len, insulin, bmi, pedigree fun, age lalu data melakukan transformasi dengan secara normalisasi, kemudian data yang ditranformasi diprediksi dengan menggunakan metode *forward selection* dan model svm. Sebelum memasuki tahap prediksi, dilakukan pengoptimalan menggunakan metode *forward selection*. Setelah melalui tahap pengoptimalan model, data pada penelitian ini diprediksi dengan model yang digunakan yaitu svm. Hasil dari model svm digeneralisasikan dengan kernel *radial basis function* (rbf). Implementasi prediksi dengan model svm ini menggunakan libary, jupyter dan bahasa pemrograman python,

### Pembangunan Perangkat Lunak

Pada tahapan ini, dengan melakukan perancangan dan pembangunan perangkat lunak untuk mempermudah melakukan prediksi dan visualisasi data yang digunakan. IDE yang akan dibangun adalah menggunakan bahasa pemrograman python dan flask sebagai mvc.

### Pengujian & Evaluasi

Dalam tahapan ini adalah mengidentifikasi hasil akurasi prediksi yang telah dilakukan oleh model svm dengan mengevaluasi hasil akurasi tersebut menggunakan pengujian *confusion matrix* serta *mean square error*.

### Pelaporan & Publikasi

Pelaporan dan publikasi merupakan tahap dimana penelitian yang telah dilakukan, didokumentasikan dan dipublikasi pada seminar nasional/internasional dalam bidang Informatika

## Sistematik Penulisan

Sistematik penulisan yang digunakan dalam penulisan tugas akhir ini adalah sebagai berikut:

BAB I : PENDAHULUAN

BAB I menjelaskan tentang latar belakang penelitian, rumusan masalah, tujuan penelitian, luaran dan manfaat penelitian, metode penelitian, dan sistematik penulisan yang digunakan dalam laporan penelitian.

BAB II : TINJAUAN PUSTAKA

BAB II menjelaskan tentang peninjauan kembali pustaka tentang masalah yang berkaitan dalam membantu menyelesaikan permasalahan pada penelitian.

BAB II I : ANALISIS DAN PERANCANGAN

BAB III menggambarkan perancangan analisis sistem berjalan dan keluaran yang dihasilkan serta perancangan perangkat lunak yang digambarakan menggunakan diagram *Unified Modeling Language*(UML), dan perancangan antarmuka.

BAB IV : IMPLEMENTASI DAN PENGUJIAN

BAB IV berisi hasil implementasi dari rancangan yang telah dilakukan pada BAB III. Bab ini juga berisi pengujian yang dilakukan adalah pengujian hasil akurasi dengan data

BAB V : KESIMPULAN DAN SARAN

BAB V berisi kesimpulan tentang hasil penelitian berdasarkan tahapan yang telah dilakukan dan saran penelitian baik berupa usulan pengembangan maupun yang lainnya agar penelitian ini dapat lebih baik lagi pada masa yang akan datang

# BAB II

# TINJAUAN PUSTAKA

Ada beberapa penelitian terdahulu yang menjadi acuan dalam penelitian ini yaitu Forward Selection pada Support Vector Machine untuk Memprediksi Kanker Payudara. Pada penelitian ini, penulis membandingkan 3 kernel pada svm yaitu rbf, polynominal, dot. Dengan menggunakan dataset kanker payudara, dimana memiliki 10 atribut yang terdiri dari 9 atribut dan 1 label. Pada 9 atribut terdiri dari *Age, BMI(Body Mass Index), Glucose, Insulin, HOMA(Homeostatis Model Assesment), Leptin, Adiponectim, Resitin,* dan *MCP-1(Chemokine Monocyte)*. Dan pada atribut label diantaranya 1 = *Healthy Control*, 2 = *Patients*. Dari dataset tersebut, penulis menggunakan 3 kernel svm tanpa *forward selection* dan dengan *forward selection*. Data dipecah dengan menerapkan *cross validation* bernilai 10 untuk membagi data menjadi 2 bagian yaitu 90% data *training* dan 10% data *testing*. Lalu penulis prediksi data tersebut tanpa *forward selection* dan hasil yang didapat untuk svm dot 72.12%, lalu polynominal 63.71%, rbf 70.98%. Kemudian melakukan ulang dengan menggunakan *forward selection* dan hasil yang didapat untuk svm dot 79.39%, polynominal 71.36%, serta rbf 85.38%. Dari hasil tersebut dapat disimpulkan bahwa metode svm rbf dengan *forward selection* menghasilkan akurasi yang tinggi dibandingkan dengan 2 kernel lainnya [18].

Pada penelitian lainnya yaitu Prediksi Harga Kedelai Lokal dan Kedelai Import dengan Metode *Support Vector Machine* berbasis *Forward Selection*. Penelitian menggunakan data harga kedelai lokal maupun kedelai import dari tanggal 3 januari 2011 sampai 10 maret 2015 yang mana data tersebut masih berupa data *univariate*. Penulis melakukan konversi menjadi data *multivariate* dalam bentuk 2 inputan, 3 inputan, 4 inputan, sampai dengan 7 inputan. Maksud dari 2 inputan yaitu melihat harga dari 2 hari sebelumnya, sedangkan 3 inputan, melihat harga dari 3 hari sebelumnya dan begitu seterusnya untuk mencari nilai rmse terendah. Serta menggunakan *cross validation* 10 dan 15, serta mengubah nilai c / cost pada dataset yaitu nilai cost = 0.1, cost = 0.3, cost = 0.5. Kemudian melakukan uji coba dataset yang telah diubah menggunakan svm dengan kernel dot, polynominal, rbf, neural, anova dan memasukan nilai cost serta range yang telah ditentukan, dan hasil yang didapatkan adalah dengan seleksi atribut menggunakan *forward* selection parameter inputan data 4 hari sebelumnya, *k-fold* = 10, nilai c = 0.1 memperoleh nilai rmse 74.276, sedangkan parameter inputan data 5 hari sebelumnya, *k-fold* = 15, nilai c = 0,1 dengan melakukan optimasi *forward selection* memperoleh nilai rmse 3.738. Dari hasil tersebut, dapat dijelaskan dengan prediksi harga kedelai lokal dan import menggunakan svm yang berbasis *forward selection* memiliki performa yang lebih baik [19]

## Data Mining

Data mining adalah proses pengumpulan sebuah informasi penting pada suatu data yang berukuran besar. Untuk pengumpulan data tesebut dapat dilakukan melalu proses perhitungan statistika, matematika, maupun teknologi AI (*Artificial Intelligent*). Istilah lain dari data mining adalah penambangan data yang berbentuk sebuah tool untuk melakukan analisa dengan teknik penyaringan informasi secara lebih akurat. Teknik tersebut biasanya dilakukan untuk menemukan beberapa pola-pola tertentu yang masih memiliki relevansi dengan *goals* [7][8]. Ada beberapa teknik yang terkenal pada data mining. Diantaranya yaitu teknik klasifikasi. Klasifikasi adalah sebuah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data dengan tujuan untuk memperkirakan kelas dari suatu objek yang labelnya tidak diketahui [6]. Hal ini juga dapat dikatakan sebagai pembelajaran (klasifikasi) yang memetakan sebuah unsur (item) data kedalam salah satu dari beberapa kelas yang sudah didefinisikan. Teknik ini banyak digunakan dalam berbagai bidang, seperti dibidang kesehatan yaitu prediksi penyakit kanker [18], mendiagnosis riwayat kesehatan pasien [20] [21].

Beberapa tahapan-tahapan yang perlu dilakukan pada data mining. Hal yang pertama adalah seleksi data, pemilihan data sebelum tahap penggalian informasi dimulai. Data hasil seleksi yang akan digunakan untuk 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 data mining, adalah proses mencari pola atau informasi dalam data terpilih dengan menggunakan teknik atau metode tertentu. Terakhir evaluasi, dapat diartikan menampilkan pola informasi yang telah diolah pada tahap data mining kedalam bentuk yang mudah dimengerti [22]. Tahapan tersebut dapat dilihat pada Gambar 2.1 Tahapan-tahapan proses dalam data mining



**Gambar 2. 1 Tahapan-tahapan proses dalam data mining**

Ada penelitian dengan menggunakan metode data mining yaitu, penelitian yang dilakukan oleh Damir Imamovic, Elmir Babovic, Nina Bijedic, pada tahun 2020. Menggunakan metode data mining pada prediksi kematian pada pasien dengan penyakit kardiovaskular menggunakan metode data mining. Dataset yang digunakan pada penelitian ini mencakup 71 atribut dan dua spesial yaitu id pasien dan atribut label status vital (12 bulan FU), yang berisi 2 kelas Hidup dan Mati. Berdasarkan penelitian yang dilakukan, menyimpulkan bahwa metode data mining akurat dalam memprediksi kematian pada pasien penyakit kardiovaskular [23].

## Diabetes Melitus

Diabetes melitus adalah penyakit kronis yang ditandai dengan tingginya kadar gula (glukosa) didalam darah. Kondisi ini juga sering disebut sebagai penyakit gula. Gula yang berada didalam darah seharusnya diserap oleh sel-sel tubuh untuk kemudian diubah menjadi energi. Insulin adalah hormon yang bertugas untuk membantu penyerapan glukosa dalam sel-sel tubuh untuk diolah menjadi energi, sekaligus menyimpan sebagian glukosa sebagai cadangan energi [24]. Diabetes dapat disebabkan oleh beberapa kondisi, seperti kurangnya produksi insulin pada pankreas, gangguan respon tubuh pada insulin, adanya pengaruh hormon lain yang menghambat kinerja insulin. Apabila kondisi ini diabaikan dan kadar gula darah dibiarkan tinggi tanpa dikendalikan, maka akan melahirkan komplikasi yang membahayakan [25]. Tingginya kadar gula dalam darah pada penyakit diabetes disebabkan oleh beberapa hal. Pertama kondisi autoimun, kondisi autoimun yang menyebabkan diabetes melitus terjadi ketika sistem kekebalan tubuh menyerang dan menghancurkan sel-sel pankreas yang bertugas menghasilkan hormon insulin. Kedua resistansi insulin, resistansi insulin sendiri membuat sel tubuh tidak bisa menerima gula darah untuk kemudian diolah menjadi energi. Hal ini memberi sinyal bahwa tubuh kekurangan darah, sehingga memecah kembali glikogen. Pada akhirnya, gula akan terus menumpuk dan menyebabkan kadar gula tinggi, atau disebut hiperglikemia. Penyakit ini dapat terdeteksi dengan pengobatan secara dini. Serta dapat dikontrol dengan mengubah pola hidup sehat. Perubahan gaya tersebut meliputi pola makan sehat dan olahraga secara teratur [26].

Pada penelitian ini, penulis yang bernama V. Anuja Kumari, R, Chita menggunakan topik memprediksi diabetes melitus pada tahun 2013. Pada topik tersebut, dataset digunakan dalam penelitian ini adalah pasien diabetes yang setidaknya berusia 21 tahun. Dari hasil penelitian ini dapat dibuktikan bahwa penyakit diabetes dapat terprediksi secara akurat dengan menambahkan metode svm [27]

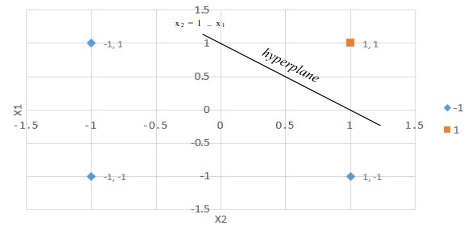
## Support Vector Machine (Radial Basis Function)

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* [13]. Rumus untuk margin.

Untuk mencari *hyperplane*, perlu meminimalkan nilai margin.

Dengan syarat untuk mencari persamaan hyperplane:

Sehingga mendapatkan *hyperplane* seperti pada Gambar 2.2 Hyperplane



**Gambar 2. 2 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 dimana 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*. Kernel RBF adalah fungsi yang nilainya tergantung pada jarak dari titik asal atau dari beberapa titik [13]. Sebagai berikut:

Dengan menggunakan jarak di ruang asli, kernel ini dapat menghitung produk titik (kesamaan) dari x dan y, maksud dari kesamaan adalah jarak sudut antara dua titik. Pada penelitian terdahulu, Liaqat Ali, Iram Wajahat, Noorbakhsh Amiri Golilarz, Fazel Keshtkar pada tahun 2020. Membahas mengenai meningkatkan prediksi karsinoma hepatoseluler melalui pengurangan dimensi dan svm yang dioptimalkan, pada penelitian ini mengusulkan sistem cerdas hybrid baru dengan membandingkan tiga algoritma, yaitu *linier discriminant analysis*(LDA), *support vector machine*(SVM), serta algoritma Genetika. Hasil eksperimen pada dataset HCC menunjukan bahwa, kompleksitas yang dihasilkan dari algoritma SVM lebih unggul dibandingkan kedua metode lain [28]

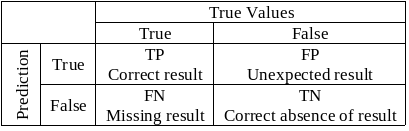
## Forward Selection

*Forward Selection* adalah salah satu prosedur bertahap yang bertujuan untuk menambah variabel yang dikendalikan satu per satu ke dalam persamaan yang didasarkan pada Alpha tertentu untuk masukan. Alpha untuk masukan merupakan nilai yang menentukan salah satu prediktor yang saat ini tidak dalam model, harus ditambahkan ke model. Nilai P dari masing-masing prediktor dalam model ini tidak dibandingkan dengan tingkat ini, jika nilai P dari prediktor kurang dari tingkat, sehingga prediktor merupakan kandidat untuk dimasukkan ke dalam model. Korelasi sederhana dapat ditentukan dengan menggunakan matriks tabel korelasi sederhana. Prosedur ini akan berakhir ketika semua variabel yang masuk ke dalam model dan memiliki nilai P kurang dari Alpha tertentu untuk masukan [17] [19].

Pada penelitian terdahulu, Elin Nurlia, Ultach Enri pada tahun 2021. Penelitian ini menggunakan metode fitur seleksi *forward selection* untuk kematian akibat gagal jantung. Dataset tersebut diperoleh dari UCI *Machine Learning Repository* yang terdiri dari 299 record data pasien dengan 13 atribut. Dengan metode *forward selection* mampu meningkatkan performa algoritma C4.5 dalam melakukan klasifikasi [29]

## Confusion Matrix

*Confusion Matrix* adalah pengukuran performa untuk masalah klasifikasi machine learning dimana keluaran dapat 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) merupakan jumlah data negatif yang terdeteksi dengan benar, sedangkan *False Positive* (FP) merupakan data negatif namun terdeteksi sebagai data positif. Sementara itu, *True Positive* (TP) merupakan data positif yang terdeteksi benar. *False Negative* (FN) merupakan kebalikan dari True Positive, sehingga data posifit, namun terdeteksi sebagai data negatif [16]. Dapat dilihat pada Gambar.



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: ((TP+TN) / (TP+FP+TN+FN))\*100

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

Rumus: (TP / (TP+FP))\*100

* Recall, menggambarkan keberhasilan model dalam menemukan kembali informasi.

Rumus: (TP / (TP+FN))\*100

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

Rumus: 2 \* *precision* \* *recall* / (*precision*+*recall*)

Ada penelitian yang menguji hasil akurasi menggunakan *confusion matrix*. Yang dilakukan oleh J. Pradeep Kandhasamy, S. balamurali. Menyatakan bahwa pada penelitian prediksi pada data pasien diabetes menggunakan metode *decision tree* J48. Dengan melakukan pengujian model menggunakan *confusion matrix* sangat efektif dalam mengevaluasi data yang diprediksi [30]

## Mean Square Error

Seperti yang diketahui bahwa, *Mean Square Error* atau disingkat MSE adalah metode peramalan untuk mengevaluasi hasil data. Metode MSE secara umum digunakan untuk mengecek estimasi berapa nilai kesalahan pada peramalan. Nilai MSE yang rendah atau nilai mendekati nol menunjukan bahwa hasil peramalan sesuai dengan data aktual dan bisa dijadikan untuk perhitungan peramalan di periode mendatang [8]. Rumus *Mean Square Error* sebagai berikut:

Dimana:

At = nilai aktual permintaan

Ft = nilai hasil peramalan

n = banyaknya data

# BAB III

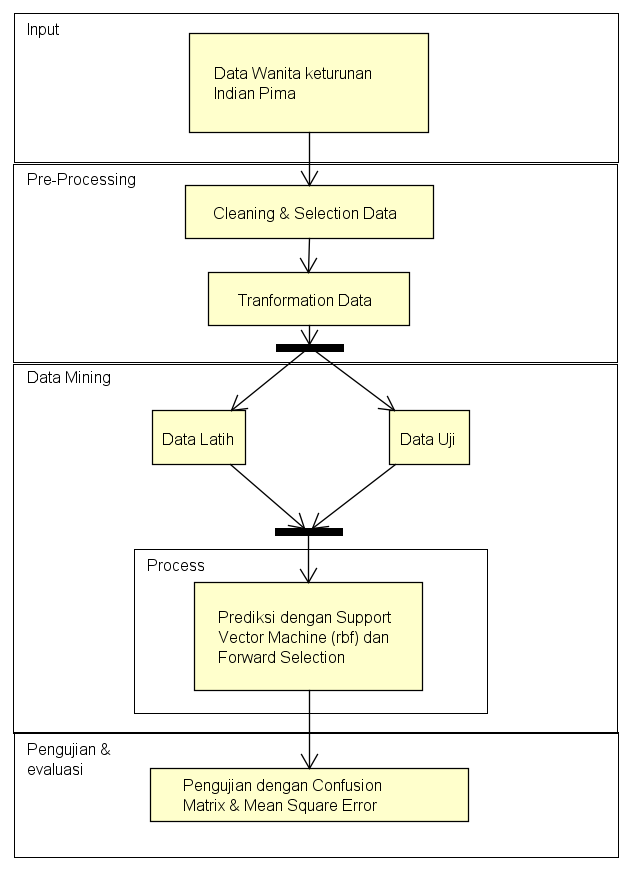
# ANALISIS DAN PERANCANGAN

## Analisis Prediksi Diabetes Menggunakan Support Vector Machine

Diabetes adalah kondisi kesehatan serius dimana gula darah penderitanya terlalu tinggi untuk diproses oleh tubuh. Gula darah tersebut akan mengendap di aliran darah dan seiring berjalannya waktu dapat menimbulkan komplikasi, seperti penyakit jantung, penyakit ginjal hingga kehilangan pengelihatan [31]. Sayangnya, masyarakat seringkali mengabaikan anjuran tersebut, sehingga tanpa disadari kadar gula yang ada pada tubuh meningkat dengan tinggi saat penyakit diabetes timbul. Hal ini menunjukan bahwa kesadaran akan bahaya penyakit diabetes yang masih rendah. Dengan mendeteksi penyakit diabetes secara dini sangatlah diperlukan dikarenakan dapat membantu penderita untuk mengontrol gula darah, mengubah pola hidup serta mencegah timbulnya komplikasi [32]. Mendeteksi penyakit diabetes dapat dilakukan dengan melakukan pemeriksaan langsung dengan dokter. Dokter akan mendiagnosis penyakit tersebut dengan melakukan wawancara medis, pemeriksaan fisik, serta pemeriksaan penunjang, seperti pemeriksaan darah, pemeriksaan berat badan, pemeriksaan insulin [33] [34].

Proses memprediksi penyakit diabetes yang dilakukan oleh sistem prediksi diabetes yang sudah ada, sedikit berbeda dengan mendeteksi yang dilakukan secara langsung dengan dokter. Jika mendeteksi secara langsung melalui dokter proses yang perlu dilakukan adalah melakukan pendaftaran dengan pihak rumah sakit, setelah itu melakukan wawancara kecil dengan dokter, pertanyaan yang terkait seperti apakah pola makan teratur, apakah sering olahraga. Setelah itu, dokter melakukan pemeriksaan kepada pasien yaitu, pemeriksaan darah, insulin, berat badan. Kemudian dokter melakukan diagnosis pada riwayat pasien apakah pasien terkena diabetes atau tidak. Sedangkan memprediksi penyakit menggunakan sistem seperti penelitian yang dilakukan oleh Prof. K. JayaMalini, Priyanka Sonar. Pada penelitian tersebut penulis menggunakan data record sebanyak 768, dengan 8 atribut dan 1 label. 8 atribut terdiri dari *total number of times pregnant, glucose/sugar level, diastolic blood pressure, body mass index, skin fold thickness in mm, insulin value in 2 hour, hereditary factor-pedigree function, age of patient in years*, dengan pembagian data 75% digunakan untuk pelatihan dan 25% digunakan untuk pengujian. Setelah itu melakukan proses prediksi menggunakan 4 algoritma yaitu *Decision Tree*, *Support Vector Machine*, *Naive Bayes*, *Artificial Neural Network*. Dari 4 metode tersebut, hasil penelitian menunjukan bahwa algoritma *Support Vector Machine* adalah algoritma yang cocok untuk memprediksi penyakit diabetes dengan hasil akurasi 82.0% menggunakan pengujian *Confusion Matrix* [35].

Dalam penelitian ini, metode yang akan digunakan dalam prediksi penyakit diabetes adalah algoritma *Support Vector Machine* (*Radial Basis Funciton*)serta dengan penambahan metode *Forward Selection* agar hasil akurasi dari algoritma *support vector machine* (rbf) lebih tinggi [19]. Pada penelitian ini data yang digunakan adalah wanita keturunan indian pima yang didapatkan pada website kaggle dengan keluarannya atau label yaitu *YES* dan *NO*. Perancangan sistem yang akan dibuat dapat dilihat pada Gambar 3.1



**Gambar 3. 1 Perancangan Sistem Prediksi Diabetes**

## Data Collection

Pada penelitian ini, data yang akan digunakan adalah data hasil pemeriksaan dengan jenis kelamin wanita keturunan indian pima dengan jumlah *record* 1017 data. Dataset yang digunakan pada penelitian ini memiliki kesamaan dalam atribut data dengan penelitian sebelumnya yaitu mengenai prediksi diabetes menggunakan metode *support vector machine*[27].

### Deskripsi Atribut Data

Sebelum memasuki tahap proses data mining, dataset ini perlu dijelaskan tiap atribut yang ada pada data tersebut, agar dapat menentukan atribut yang digunakan dari data pasien ini. Deskripsi atribut dapat dilihat pada Tabel 3.1.

**Tabel 3. 1 Deskripsi Atribut Data**

|  |  |  |
| --- | --- | --- |
| **No** | **Atribut Data** | **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\_Fun | 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 |

Deskripsi atribut data pada tabel diatas menjelaskan mengenai keterangan yang ada pada atribut data, dengan memiliki 9 atribut data yang digunakan pada penelitian ini.

### Data Pasien Wanita Keturunan Indian Pima

Berikut adalah tampilan data pasien wanita keturunan indian pima yang dapat dilihat pada Tabel 3.2.

**Tabel 3. 2 Contoh Data Pemeriksaan Pasien Wanita Keturunan Indian Pima**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **N\_pregnant** | **Glucose\_conc** | **BP** | **Skin\_Len** | **Insulin** | **BMI** | **Pedigree\_Fun** | **Age** | **Output** |
| 6 | 148 | 72 | 35 | 0 | 33.6 | 0.627 | 50 | YES |
| 1 | 85 | 66 | 29 | 0 | 26.6 | 0.351 | 31 | NO |
| 8 | 183 | 64 | 0 | 0 | 23.3 | 0.672 | 32 | YES |
| 1 | 89 | 66 | 23 | 94 | 28.1 | 0.167 | 21 | NO |
| 0 | 137 | 40 | 35 | 168 | 43.1 | 2.288 | 33 | YES |
| 5 | 116 | 74 | 0 | 0 | 25.6 | 0.201 | 30 | NO |
| 3 | 78 | 50 | 32 | 88 | 31 | 0.248 | 26 | YES |
| 10 | 115 | 0 | 0 | 0 | 35.3 | 0.134 | 29 | NO |
| 2 | 197 | 70 | 45 | 543 | 30.5 | 0.158 | 53 | YES |
| 8 | 125 | 96 | 0 | 0 | 0 | 0.232 | 54 | YES |
| 4 | 110 | 92 | 0 | 0 | 37.6 | 0.191 | 30 | NO |
| 10 | 168 | 74 | 0 | 0 | 38 | 0.537 | 34 | YES |
| 10 | 139 | 80 | 0 | 0 | 27.1 | 1.441 | 57 | NO |
| 1 | 189 | 60 | 23 | 846 | 30.1 | 0.398 | 59 | YES |
| 5 | 166 | 72 | 19 | 175 | 25.8 | 0.587 | 51 | YES |

## Pre Processing

Tahap *pre-processing* adalah tahapan untuk merubah data menjadi data yang siap diuji. Pada tahap *pre-processing*, data diproses melalui tahap data *cleaning*, *selection*, serta data *transformation*.

### Data Cleaning & Selection

Pada tahap ini dilakukan pemilihan himpunan data, menciptakan himpunan data target, dan memfokuskan pada subset variabel. Pada tahap ini pula dilakukan perbaikan data serta pembuangan data yang tidak konsisten dan mengandung noise [36].

Dalam penelitian ini, dataset yang digunakan akan dilakukan proses *cleaning* dan *selection* data untuk menentukan atribut yang akan digunakan dalam sistem. Dari proses tersebut, atribut data yang digunakan akan melakukan *selection* dan *cleaning*, atribut yang dipilih meliputi *glucose\_conc*, *bp*, *insulin*, *bmi*, *skin\_len*[]. Hal tersebut dikarenakan atribut yang dipilih memiliki variabel 0 terbanyak, oleh sebab itu maka dilakukan pergantian pada variabel 0 menjadi nilai nan dengan menghitung secara total, kemudian tiap atribut dihitung secara median berdasarkan atribut output. Berikut Tabel 3.3 dibawah yang akan digunakan.

**Tabel 3. 3 Data Pasien Wanita Indian Pima**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **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 | YES |
| 1 | 85.0 | 66.0 | 29.0 | 95.5 | 26.6 | 0.351 | 31 | NO |
| 8 | 183.0 | 64.0 | 32.0 | 168.5 | 23.3 | 0.672 | 32 | YES |
| 1 | 89.0 | 66.0 | 27.0 | 94.0 | 28.1 | 0.167 | 21 | NO |
| 0 | 137.0 | 75.5 | 35.0 | 168.0 | 43.1 | 0.467 | 33 | YES |
| 5 | 116.0 | 74.0 | 27.0 | 95.5 | 25.6 | 0.201 | 30 | NO |
| 3 | 78.0 | 50.0 | 32.0 | 88.0 | 31.0 | 0.248 | 26 | YES |
| 10 | 115.0 | 70.0 | 27.0 | 95.5 | 35.3 | 0.134 | 29 | NO |
| 2 | 197.0 | 70.0 | 32.0 | 543.0 | 30.5 | 0.158 | 53 | YES |

Dapat dilihat pada tabel diatas adalah dataset hasil dari processing. Atribut pada dataset yang di *selection* yaitu *n\_pregnant*, *glucose\_conc*, *bp*, *skin\_len*, *insulin*, *bmi*, *pedigree\_fun*, *age*.Pada atribut yang dipilih dilakukan perhitungan dengan rumus median untuk mencari variabel tiap atribut yang kemudian nilai tersebut dilakukan proses *filtering*,dimana variabel diproses melalui *filtering* berdasarkan hasil dari atribut *output*. Lalu melakukan *filtering* semua atribut.

### Data Tranformasi

Pada Tahap ini dilakukan *transformasi* data yang telah melalui proses *cleaning* & *selection*. Tahapan ini mengubah variabel data menjadi bentuk data seperti numerik, nominal, dll. Tahap *transformasi* ini menggunakan proses normalisasi dengan mengganti data nominal menjadi data numerik.

**Tabel 3. 4 Contoh Transformation Dataset**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **N\_pregnant** | **Glucose\_conc** | **BP** | **Skin\_Len** | **Insulin** | **BMI** | **Pedigree\_Fun** | **Age** | **Output** |
| 6 | 148 | 72 | 35 | 0 | 33.6 | 0.627 | 50 | 1 |
| 1 | 85 | 66 | 29 | 0 | 26.6 | 0.351 | 31 | 0 |
| 8 | 183 | 64 | 0 | 0 | 23.3 | 0.672 | 32 | 1 |
| 1 | 89 | 66 | 23 | 94 | 28.1 | 0.167 | 21 | 0 |
| 0 | 137 | 40 | 35 | 168 | 43.1 | 2.288 | 33 | 1 |
| 5 | 116 | 74 | 0 | 0 | 25.6 | 0.201 | 30 | 0 |
| 3 | 78 | 50 | 32 | 88 | 31 | 0.248 | 26 | 1 |
| 10 | 115 | 0 | 0 | 0 | 35.3 | 0.134 | 29 | 0 |
| 2 | 197 | 70 | 45 | 543 | 30.5 | 0.158 | 53 | 1 |
| 8 | 125 | 96 | 0 | 0 | 0 | 0.232 | 54 | 1 |
| 4 | 110 | 92 | 0 | 0 | 37.6 | 0.191 | 30 | 0 |
| 10 | 168 | 74 | 0 | 0 | 38 | 0.537 | 34 | 1 |
| 10 | 139 | 80 | 0 | 0 | 27.1 | 1.441 | 57 | 0 |
| 1 | 189 | 60 | 23 | 846 | 30.1 | 0.398 | 59 | 1 |
| 5 | 166 | 72 | 19 | 175 | 25.8 | 0.587 | 51 | 1 |

Dari data diatas adalah hasil dari tranformasi data, dimana dataset yang digunakan dalam penelitian ini memiliki data campuran yaitu 8 atribut yaitu *n\_pregnant*, *glucose\_conc*, *bp*(*bloodpressure*), *skin\_len*, *insulin*, *bmi*(*body mass index*), *pedigree\_fun*, serta *age*, dimana 8 atribut pada dataset yang digunakan memiliki jenis data numerik sedangkan pada atribut *output* memiliki jenis data nominal. Maka penelitian ini mengubah data pada atribut *output* menjadi numerik dengan menggolongkan menjadi 2 golongan yaitu 1 menandakan pasien terkena penyakit diabetes, lalu 0 menandakan pasien tidak terkena diabetes.

## Proses Prediksi Diabetes

Sebelum memasuki tahap prediksi, data yang telah diproses melalui tahapan *cleaning*, *selection*, serta *tranformation*. Data yang digunakan melakukan pemisahan data atau *spliting data*. Data dibagi menjadi 2 bagian yaitu data latih dan data uji. Untuk presentase pada dataset penelitian ini, adalah 80% digunakan sebagai data latih, sedangkan 20% digunakan sebagai data uji.

Lalu masuk ke tahapan prediksi, dimana dataset melakukan optimalisasi dengan metode *forward selection* sebelum diprediksi oleh algoritma *support vector machine* (rbf)[].

|  |
| --- |
| Optimalisasi menggunakan Forward Selection |
| Algoritma |
| #melakukan pengoptimalan  sfs = SFS(SVC(), k\_features=1, forward=True,floating=False, scoring=’accuracy’, cv=10)  sfs.fit(X, y)  x\_train, x\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=5) |

Setelah melakukan optimalisasi dengan metode *forward selection*. Maka hasil yang dioptimalisasi masuk ke tahap prediksi dengan menggunakan algoritma *support vector machine* (rbf). Algoritma *support vector machine* mencari *hyperplane* pada dataset yang digunakan lalu ditingkatkan akurasi yang dihasilkan kedalam ruang dimensi yang lebih tinggi dengan menggunakan kernel *radial basis function* atau *gaussian*.

|  |
| --- |
| Prediksi menggunakan svm (rbf) |
| Algoritma |
| #mencari akurasi kernel terbaik  for k in(‘linear’, ‘rbf’, ‘poly’, ‘sigmoid’):  model = SVC(kernel=k)  model.fit(x\_train, y\_train)  y\_pred = model.predict(x\_test)  score = accuracy\_score(y\_test, y\_pred)\*100  print(f”Akurasi skor dari {k} adalah {round(score, 1}%”)  #menggunakan kernel rbf  model SVC(kernel=’rbf’)  model.fit(x\_train, y\_train)  #prediksi data latih  y\_pred = model.predict(x\_train)  #prediksi data uji  y\_pred = model.predict(x\_test) |

Pada code yang diatas adalah rancangan yang akan digunakan dalam implementasi prediksi penyakit diabetes menggunakan algoritma *Support Vector Machine* (RBF) berbasis metode *Forward Selection*.

## Perancangan Perangkat Lunak

Tahapan ini, merupakan tahapan untuk merancang perangkat lunak yang akan dibangun dengan *Unified Modeling Language* (UML).

### Analisis Kebutuhan Perangkat Lunak

Pada tahapan ini, merupakan analisis kebutuhan perangkat lunak yang akan dibangun yaitu user melakukan *pre processing* pada dataset dengan klik *button process data* dan menampilkan hasil processing, user melakukan prediksi pada data yang telah dipisah yaitu data latih & data uji, serta user menampilkan evaluasi dari hasil akurasi yang telah diprediksi.

### Deskripsi Aktor Sistem Prediksi

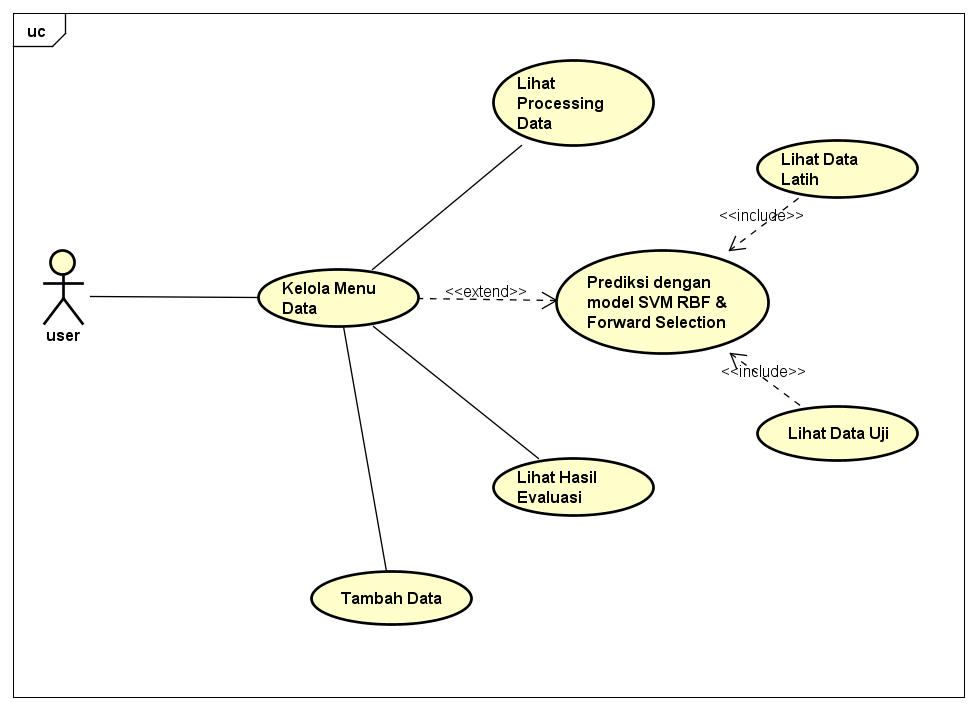
Deskripsi aktor yaitu menjelaskan definisi aktor yang terlibat yaitu user. Penjelasan mengenai deskripsi aktor dapat dilihat pada Tabel 3.5 Deskripsi Aktor.

**Tabel 3. 5 Deskrips Aktor**

|  |  |  |
| --- | --- | --- |
| **No** | **Aktor** | **Deskripsi** |
| 1 | User | Merupakan aktor yang berinteraksi dengan halaman website. |

### Use Case Diagram

*Use Case diagram* adalah diagram yang menggambarkan aktor (*user*) yang berinteraksi dengan sistem. Use case diagram dalam penelitian ini dibangun berdasarkan analisis sistem prediksi diabetes yang telah dibuat sebelumnya. Use case diagram dapat dilihat pada Gambar 3.2 Use Case Prediksi Diabetes



**Gambar 3. 2 Use Case Prediksi Diabetes**

### Deskripsi Use Case

Berdasarkan *use case diagram* yang telah dibuat pada penelitian ini, lalu akan mendeskripsikan dalam tabel berikut pada setiap use case. *Use case* dapat dilihat pada Tabel 3.6 Deskripsi Use Case

**Tabel 3. 6 Deskripsi Use Case**

|  |  |  |
| --- | --- | --- |
| **No** | **Use Case** | **Deskripsi** |
| 1 | Kelola Menu Data | Pada use case ini, user dapat melakukan kelola menu data berdasarkan data pasien diabetes |
| 2 | Lihat Processing Data | User dapat melakukan processing pada data yang digunakan dan menampilkan hasil processing. |
| 3 | Prediksi dengan model SVM RBF & Forward Selection | User dapat melakukan prediksi pada data yang telah diprocessing dengan menentukan terkena diabetes atau tidak. |
| 4 | Lihat Data Latih | User dapat melihat hasil prediksi yang telah dilakukan oleh model yang digunakan pada data latih |
| 5 | Lihat Data Uji | User dapat melihat hasil prediksi yang telah dilakukan oleh model yang digunakan pada data uji |
| 6 | Lihat Hasil Evaluasi | User dapat melihat hasil akurasi yang didapatkan dari model yang digunakan |
| 7 | Tambah Data | User dapat menambahkan data numerik yang sesuai dengan atribut. |

### Skenario Use Case

Skenario use case menggambarkan masing-masing *use case* yang terdapat pada use case diagram. Skenario use case menggambarkan interaksi aktor(*user*) dengan *use case* berupa tanggapan dari reaksi sistem atas suatu aksi yang dilakukan oleh aktor(*user*) pada sistem yang akan dibangun. Setiap use case akan diberikan skenario yang menjelaskan secara detail yang ada pada tabel berikut.

1. **Use Case Kelola Menu Data**
2. Lihat Processing Data

Skenario use case lihat processing data menjelaskan reaksi sistem dengan memprocess data dan menampilkan data tersebut. Dapat dilihat pada Tabel 3.7 Skenario Lihat Processing Data

**Tabel 3. 7 Skenario Lihat Processing Data**

|  |  |
| --- | --- |
| Name | Lihat Processing Data |
| ID | PD001 |
| Description | Sistem dapat memungkinkan user dapat melakukan lihat hasil processing data |
| Actor | User |
| Triggres | User melihat hasil processing data |
| Pre-Condition | User akan masuk halaman processing data dan memilih button process data |
| Post-Condition | Sistem akan menampilkan hasil processing |
| **Main Scenario** | |
| **Aksi Actor** | **Reaksi Sistem** |
| 1. Memilih halaman Processing Data |  |
|  | 1. Menampilkan halaman Processing Data |
| 1. Klik button process data |  |
|  | 1. Memprocess data |
|  | 1. Menampilkan data hasil processing |

1. Lihat Data Latih

Skenario use case lihat data latih menjelaskan reaksi sistem dalam memprediksi data latih menggunakan model yang digunakan. Dapat dilihat pada Tabel 3.8 Skenario Lihat Data Latih

**Tabel 3. 8 Skenario Lihat Data Latih**

|  |  |
| --- | --- |
| Name | Lihat Data Latih |
| ID | PD002 |
| Description | Sistem dapat memungkinkan user dapat melakukan lihat hasil prediksi dari data latih |
| Actor | User |
| Triggres | User melihat hasil prediksi data latih |
| Pre-Condition | User akan masuk halaman pengujian data dan memilih button train data |
| Post-Condition | Sistem akan menampilkan hasil prediksi dari data latih |
| **Main Scenario** | |
| **Aksi Actor** | **Reaksi Sistem** |
| 1. Memilih halaman Pengujian Data |  |
|  | 1. Menampilkan halaman Pengujian Data |
| 1. Klik button Train Data |  |
|  | 1. Memprediksi data latih dengan model yang digunakan |
|  | 1. Menampilkan hasil prediksi data latih |

1. Lihat Data Uji

Skenario use case lihat data latih menjelaskan reaksi sistem dalam memprediksi data uji menggunakan model yang digunakan. Dapat dilihat pada Tabel 3.9 Skenario Lihat Data Uji

**Tabel 3. 9 Skenario Lihat Data Uji**

|  |  |
| --- | --- |
| Name | Lihat Data Uji |
| ID | PD003 |
| Description | Sistem dapat memungkinkan user dapat melakukan lihat hasil prediksi dari data uji |
| Actor | User |
| Triggres | User melihat hasil prediksi data latih |
| Pre-Condition | User akan masuk halaman pengujian data dan memilih button test data |
| Post-Condition | Sistem akan menampilkan hasil prediksi dari data uji |
| **Main Scenario** | |
| **Aksi Actor** | **Reaksi Sistem** |
| 1. Memilih halaman Pengujian Data |  |
|  | 1. Menampilkan halaman Pengujian Data |
| 1. Klik button Test Data |  |
|  | 1. Memprediksi data uji dengan model yang digunakan |
|  | 1. Menampilkan hasil prediksi data uji |

1. Lihat Hasil Evaluasi

Skenario use case lihat hasil evaluasi menjelaskan reaksi sistem dalam menghitung hasil akurasi yang didapatkan dalam memprediksi data dengan model yang digunakan. Dapat dillihat pada Tabel 3.10 Skenario Lihat Hasil Evaluasi

**Tabel 3. 10 Skenario Lihat Hasil Evaluasi**

|  |  |
| --- | --- |
| Name | Lihat Hasil Evaluasi |
| ID | PD004 |
| Description | Sistem dapat memungkinkan user dapat melakukan lihat hasil akurasi berdasarkan model yang digunakan |
| Actor | User |
| Triggres | User melihat hasil evaluasi |
| Pre-Condition | User akan masuk halaman evaluasi data dan memilih button hasil evaluasi |
| Post-Condition | Sistem akan menampilkan hasil akurasi dari evaluasi model |
| **Main Scenario** | |
| **Aksi Actor** | **Reaksi Sistem** |
| 1. Memilih halaman Evaluasi Data |  |
|  | 1. Menampilkan halaman Evaluasi Data |
| 1. Klik button Hasil Evaluasi |  |
|  | 1. Menghitung akurasi dari model yang digunakan |
|  | 1. Menampilkan hasil akurasi dari model yang digunakan |

1. Tambah Data

Skenario use case tambah data menjelaskan reaksi sistem menambahkan data ke dalam dataset yang digunakan. Dapat dillihat pada Tabel 3.11 Skenario Tambah Data

**Tabel 3. 11 Skenario Tambah Data**

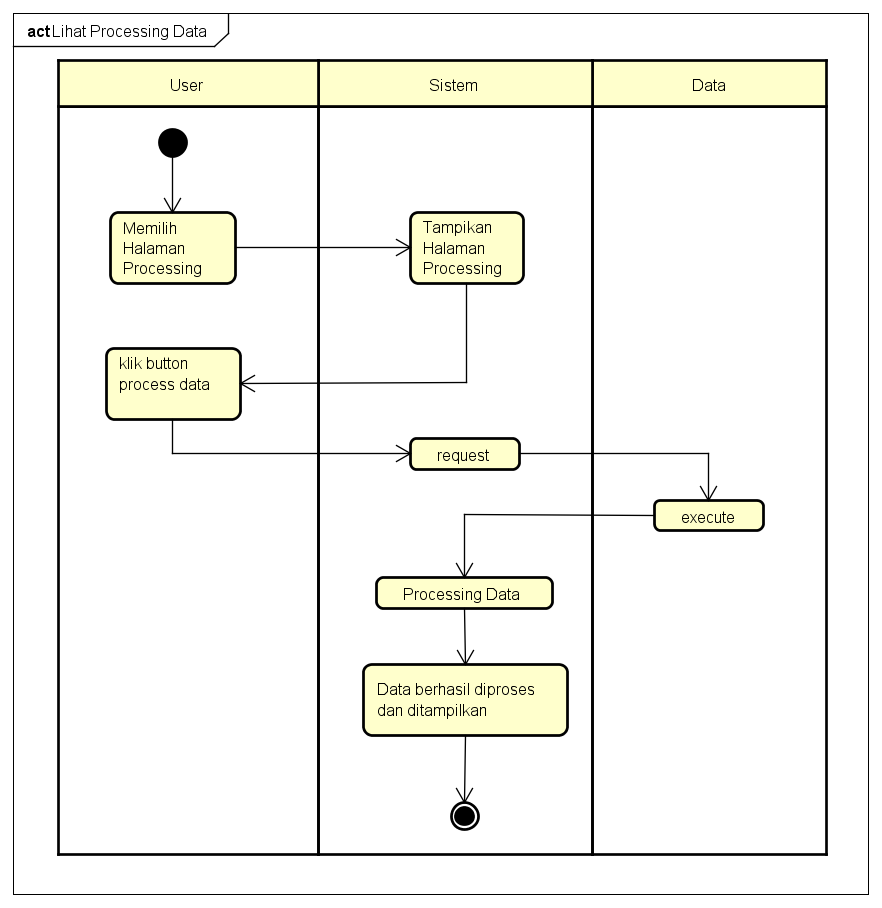
|  |  |
| --- | --- |
| Name | Tambah Data |
| ID | PD005 |
| Description | Sistem dapat memungkinkan untuk user dapat melakukan tambah data ke dataset yang digunakan |
| Actor | User |
| Triggres | User menambahkan data |
| Pre-Condition | User akan masuk halaman tambah data dan menginput data yang akan ditambahkan |
| Post-Condition | Sistem akan menampilkan notifikasi berhasil ditambahkan |
| **Main Scenario** | |
| **Aksi Actor** | **Reaksi Sistem** |
| 1. Memilih halaman Tambah Dat |  |
|  | 1. Menampilkan halaman Tambah Data |
| 1. Melakukan input data |  |
| 1. Klik button simpan |  |
|  | 1. Menambahkan data dan disimpan didataset |
|  | 1. Menampilkan notifikasi berhasil disimpan |

### Activity Diagram

*Activity Diagram* merupakan diagram yang menggambarkan alur kerja atau aktivitas dari sebuah sistem yang sedang dirancang dan bagaimana proses aliran berawal, *decision* atau keputusan yang akan terjadi, hingga bagaimana aktivitas sistem tersebut berakhir. Berikut *activity diagram* yang dirancang pada gambar berikut.

1. **Lihat Processing Data**

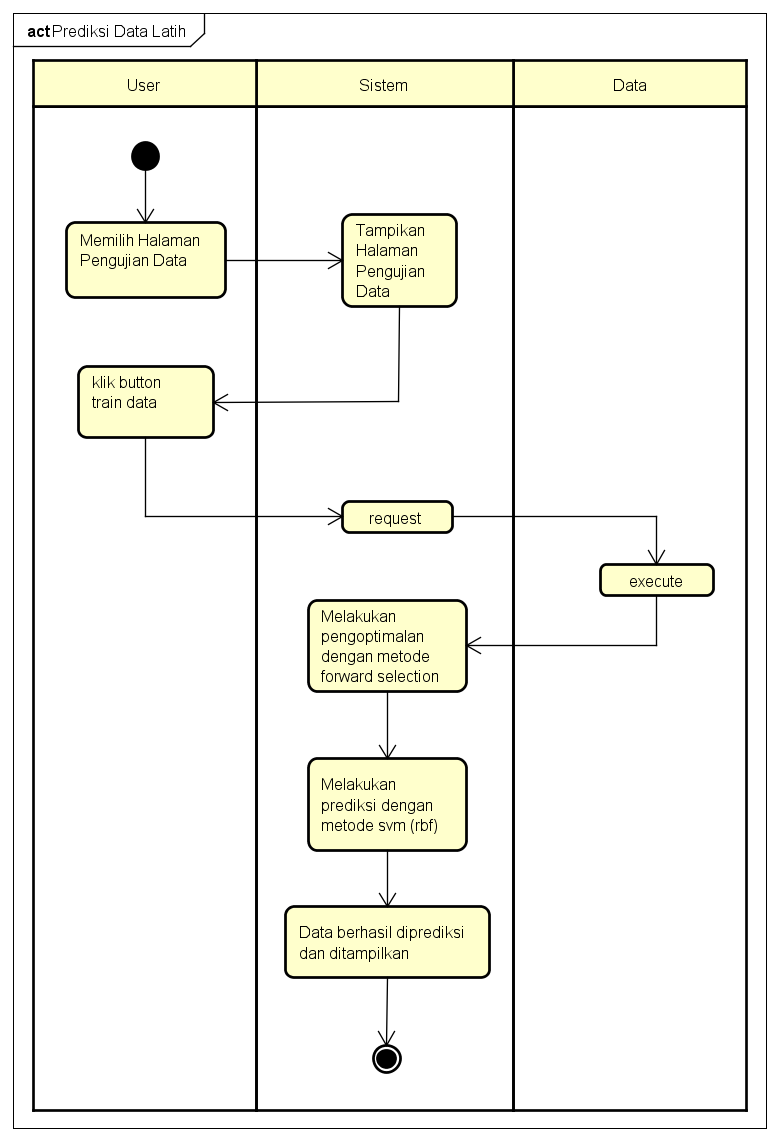
Diagram ini dimulai dari *user* memilih halaman processing data, sistem akan menampilkan halaman processing data, kemudian *user* memilih button process data, sistem akan mengambil data untuk diprocessing, kemudian sistem menampilkan hasil processing. Dapat dilihat pada Gambar 3.3 Activity Lihat Processing Data



**Gambar 3. 3 Activity Lihat Processing Data**

1. **Lihat Data Latih**

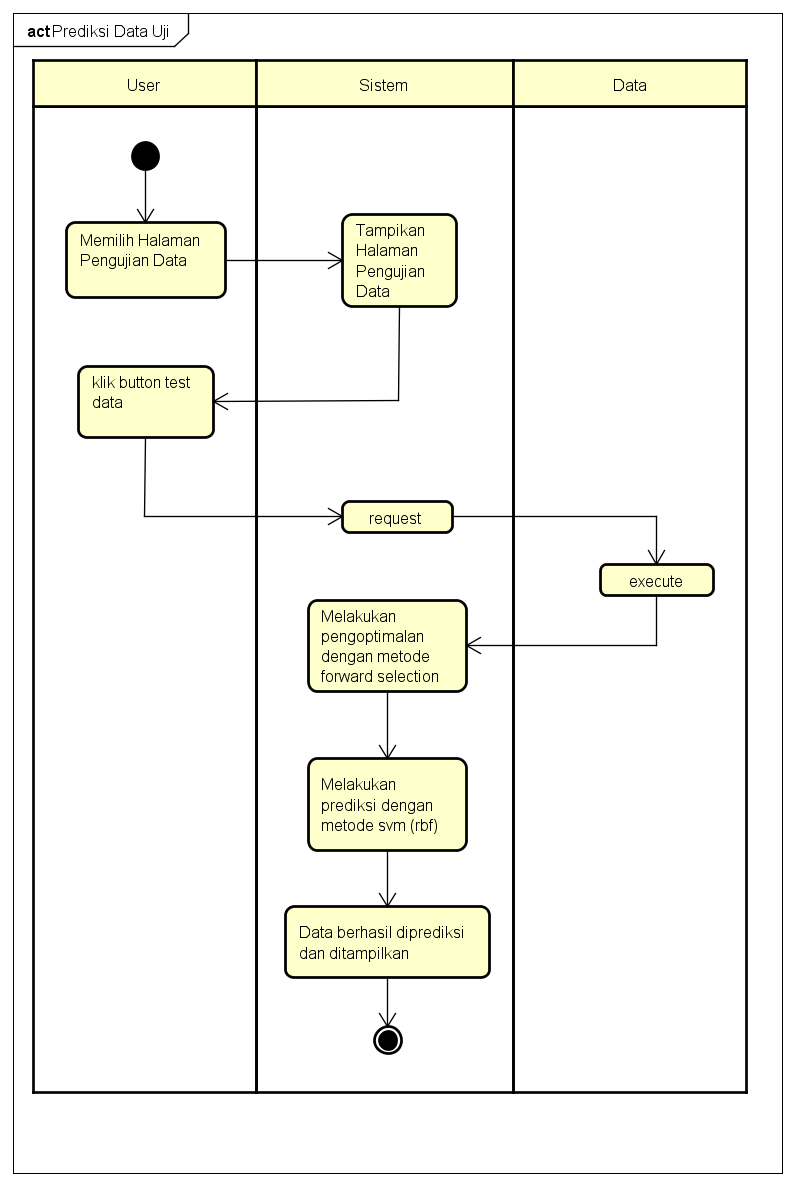
Diagram ini dimulai dari *user* memilih halaman pengujian data, sistem menampilkan halaman pengujian data, kemudian *user* memilih button *Train Data*, sistem mengambil data, lalu sistem melakukan pengoptimalan pada data dan memprediksi data yang diambil dengan model yang digunakan, setelah itu menampilkan hasil prediksi. Dapat dilihat pada Gambar 3.4 Activity Lihat Data Latih



**Gambar 3. 4 Activity Lihat Data Latih**

1. **Lihat Data Uji**

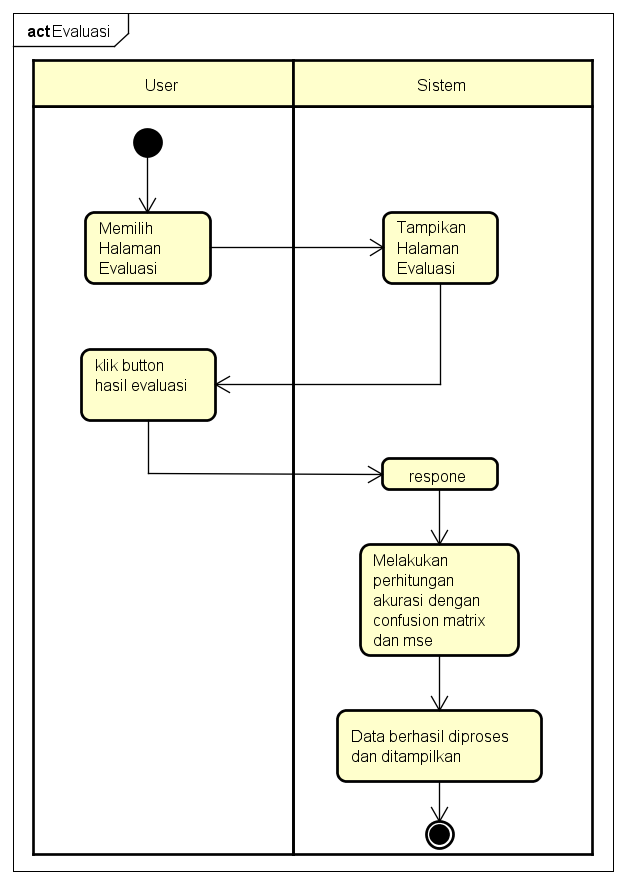
Diagram ini dimulai dari *user* memilih halaman pengujian data, sistem menampilkan halaman pengujian data, kemudian *user* memilih button *Test Data*, sistem mengambil data, lalu sistem melakukan pengoptimalan pada data dan memprediksi data yang diambil dengan model yang digunakan, setelah itu menampilkan hasil prediksi. Dapat dilihat pada Gambar 3.5 Activity Lihat Data Uji



**Gambar 3. 5 Activity Lihat Data Uji**

1. **Lihat Hasil Evaluasi**

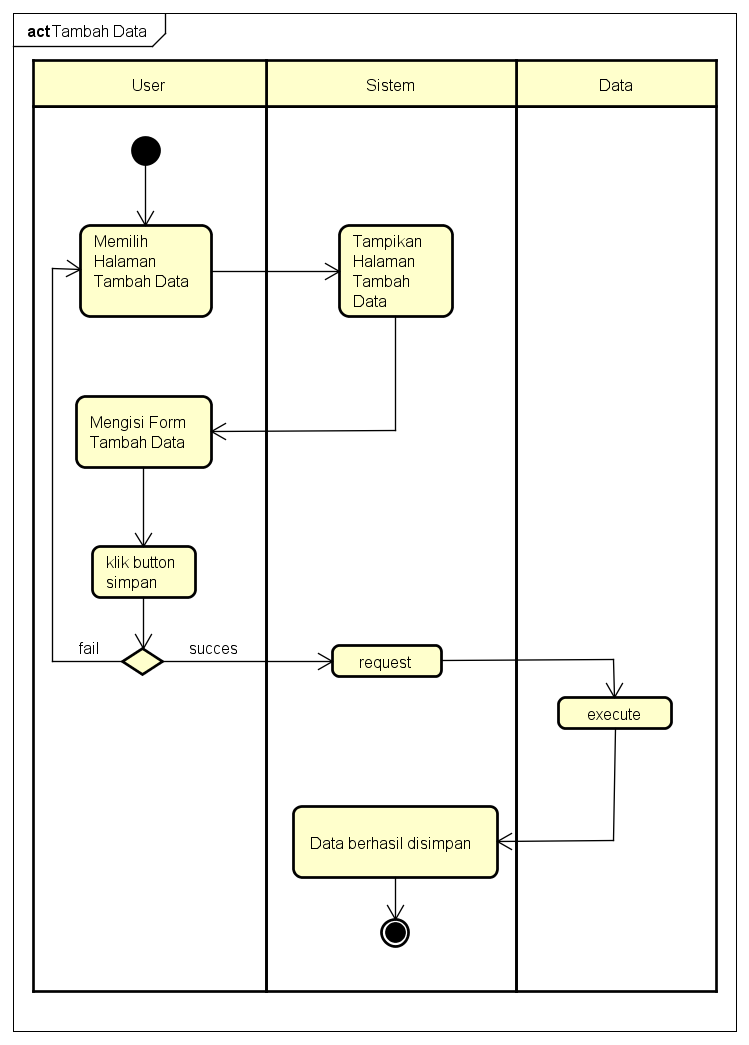
Diagram ini dimulai dari *user* memilih halaman evaluasi data, sistem menampilkan halaman evaluasi, *user* memilih button hasil evaluasi, sistem menghitung hasil akurasi yang didapatkan dari model yang digunakan. Dapat dilihat pada Gambar 3.6 Activity Lihat Hasil Evaluasi



**Gambar 3. 6 Activity Lihat Hasil Evaluasi**

1. **Tambah Data**

Diagram ini dimulai dari *user* memilih halaman tambah data, sistem menampilkan halaman tambah data, *user* mengisi form tambah data dan klik button simpan, apabila input data gagal maka akan menampilkan halaman data, jika berhasil maka menampilkan notifikasi sukses, sistem menambahkan data. Dapat dilihat pada Gambar 3.7 Activity Tambah Data



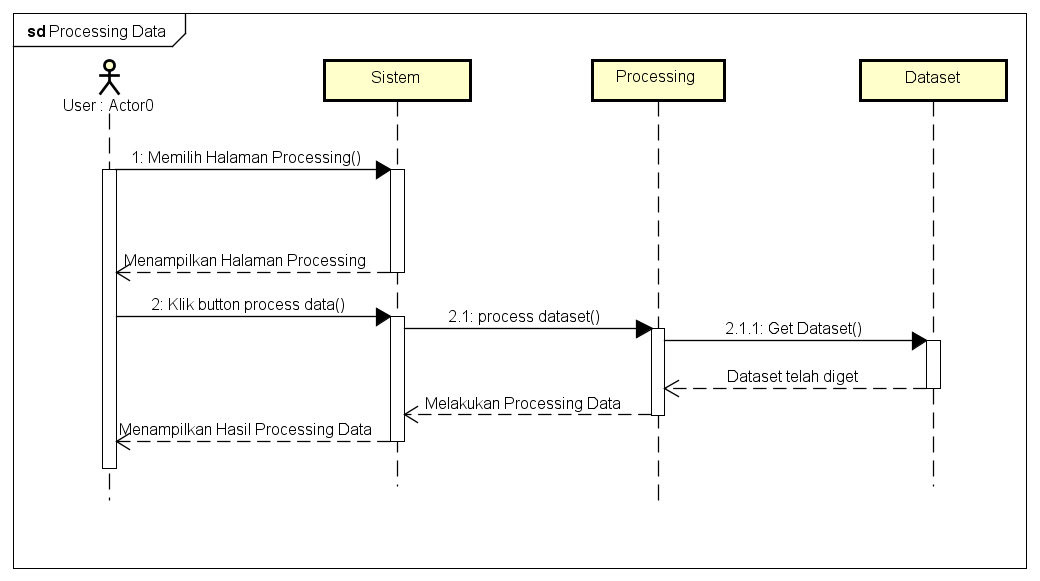
**Gambar 3. 7 Activity Tambah Data**

### Sequence Diagram

*Sequence Diagram* adalah diagram yang menggambarkan interaksi antar objek di dalam sistem yang dirancang atau aliran fungsionalitas berupa *message* yang digambarkan terhadap waktu. *Sequence Diagram* terdiri dari dimensi horizontal(objek-objek) dan dimensi vertikal(waktu). *Sequence Diagram* digunakan untuk menggambarkan skenario atau rangkaian langkah-langkah yang dilakukan sebagai respon dari sebuah *event* untuk menghasilkan output tertentu yang diawali dari apa yang mentrigger suatu aktivitas proses dan ouput yang dihasilkan.

1. **Lihat Processing Data**

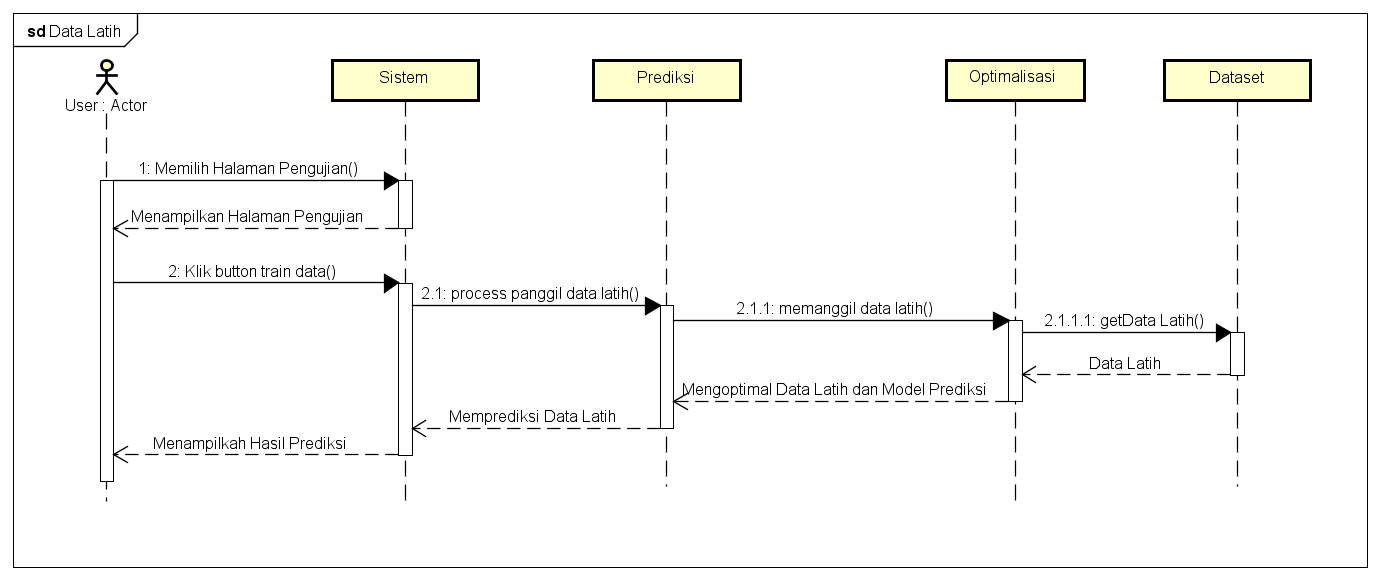
*Sequence* ini menggambarkan interaksi antar objek sebagai berikut lihat *processing* data, *user* memilih halaman *processing*, lalu menampilkan halaman *processing*, *user* memilih button process data, process dataset, dataset diambil, kemudian melakukan *processing* data, dan menampilkan hasil *processing*. Dapat dilihat pada Gambar 3.8 Sequence Processing Data



**Gambar 3. 8 Sequence Processing Data**

1. **Lihat Data Latih**

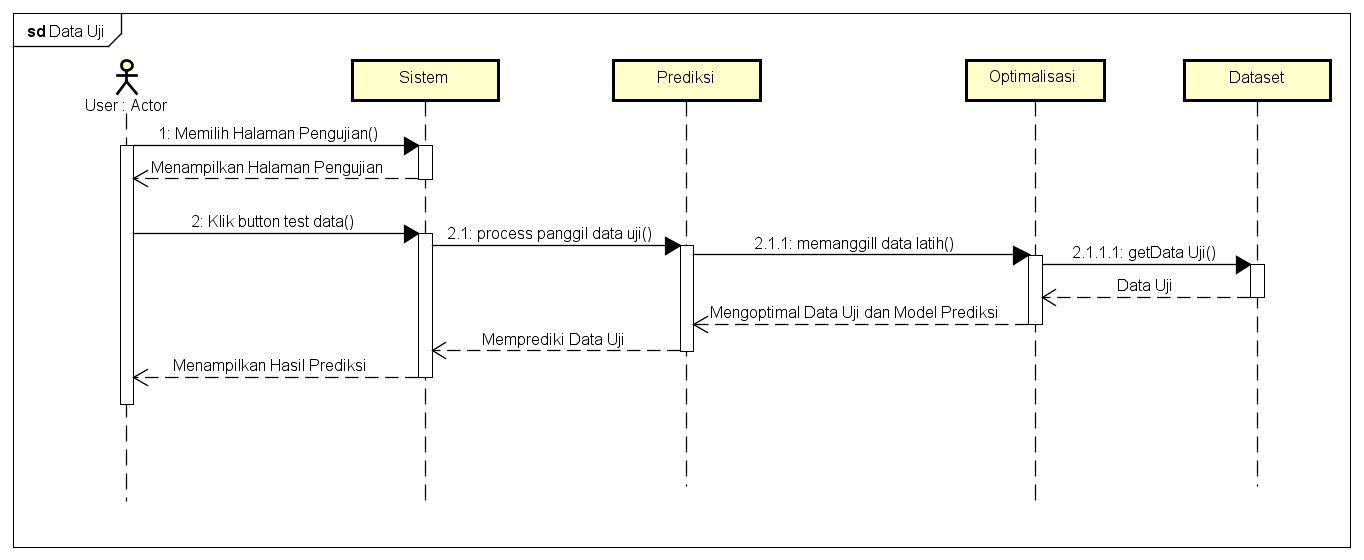
*Sequence* ini menggambarkan interaksi antar objek yaitu lihat data latih. *User* memilih halaman pengujian, lalu menampilkan halaman pengujian, *user* memilih button *train data*, sistem merespon dan memanggil data latih, mengambil data latih, mengoptimal data latih dan model prediksi, kemudian melakukan prediksi pada data latih, setelah itu menampilkan hasil prediksi. Dapat dilihat pada Gambar 3.9 Sequence Data Latih



**Gambar 3. 9 Sequence Data Latih**

1. **Lihat Data Uji**

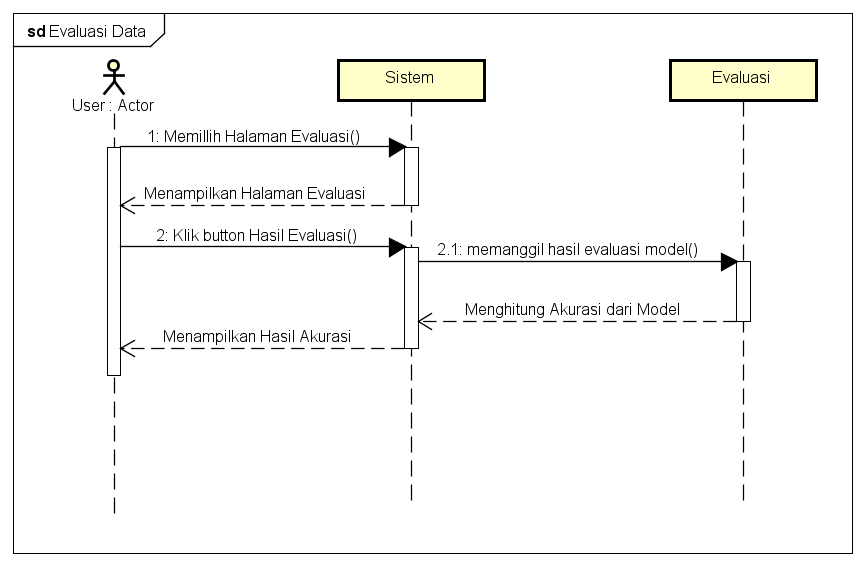
*Sequence* ini menggambarkan interaksi antar objek yaitu lihat data uji. *User* memilih halaman pengujian, lalu menampilkan halaman pengujian, *user* memilih button *test data*, sistem merespon dan memanggil data uji, mengambil data uji, lau mengoptimal data uji dan model prediksi, kemudian melakukan prediksi pada data uji, setelah itu menampilkan hasil prediksi. Dapat dilihat pada Gambar 3.10 Sequence Data Uji



**Gambar 3. 10 Sequence Data Uji**

1. **Hasil Evaluasi**

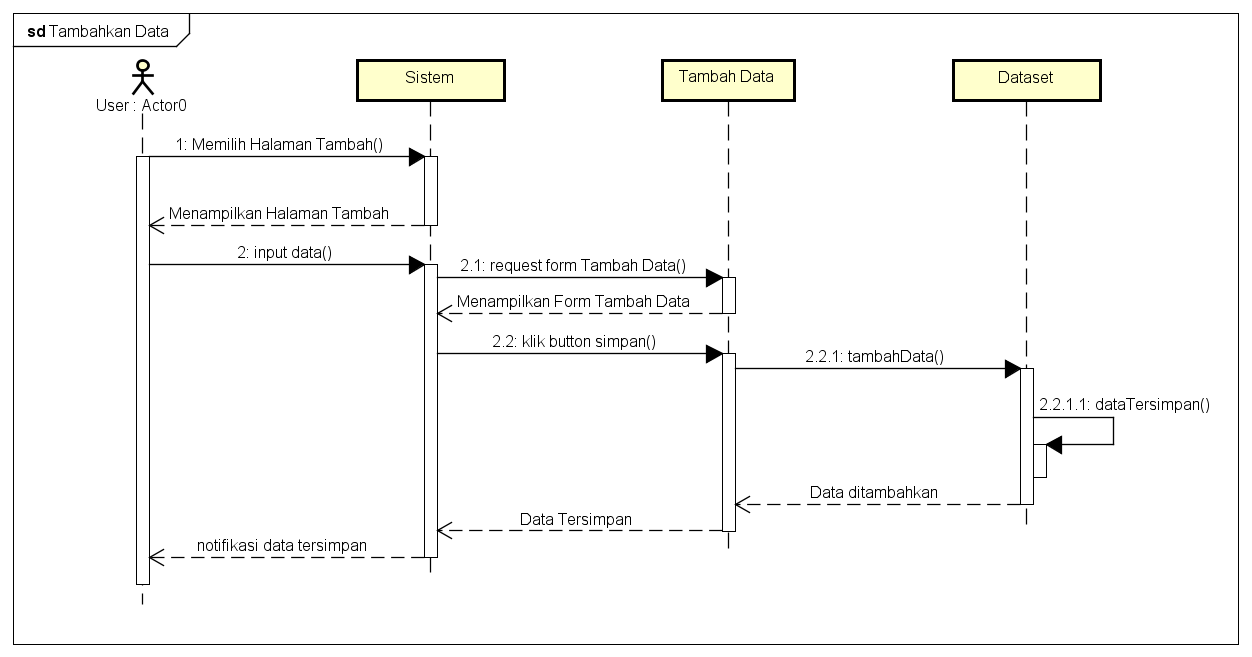
*Sequence* ini menggambarkan interaksi antar objek yaitu lihat hasil evaluasi. *User* memilih halaman evaluasi, lalu menampilkan halaman evaluasi, *user* memilih button hasil evaluasi, sistem merespon dan memanggil hasil evaluasi model, menghitung hasil akurasi dari model, kemudian menampilkan hasil akurasi. Dapat dilihat pada Gambar 3.11 Sequence Data Uji



**Gambar 3. 11 Sequence Evaluasi**

1. **Sequence Tambah Data**

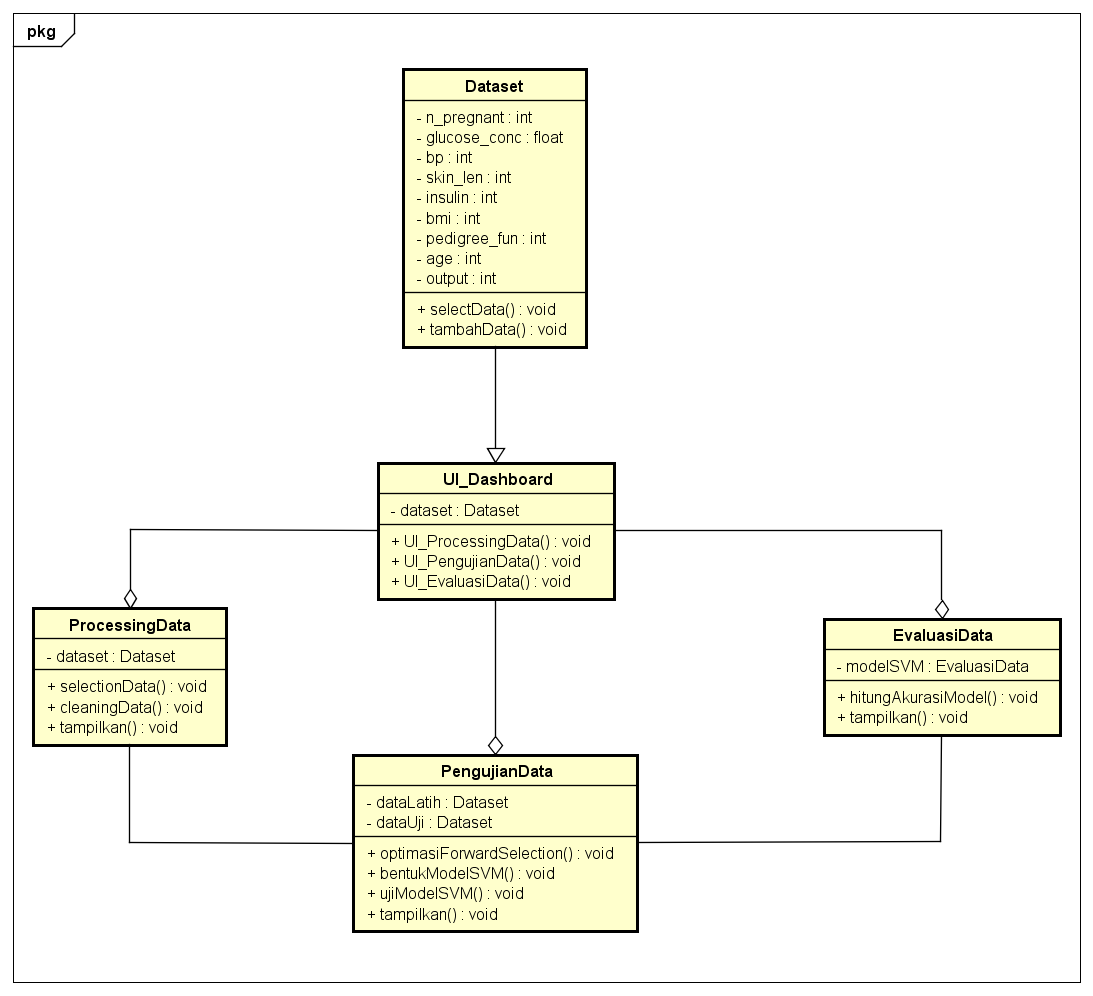
*Sequence* ini menggambarkan interaksi antar objek yaitu tambah data. *User* memilih halaman tambah data, lalu menampilkan halaman tambah data, *user* mengisi form tambah data dan klik button simpan, apabila input data gagal maka akan menampilkan halaman data, jika berhasil maka menampilkan notifikasi sukses, sistem menambahkan data. Dapat dilihat pada Gambar 3.12 Sequence Data Uji



**Gambar 3. 12 Sequence Tambah Data**

### Class Diagram

*Class Diagram* berfungsi untuk menggambarkan kelas yang terdapat pada sistem, dimana sistem yang dibangun memiliki beberapa kelas yang saling terhubung satu sama lain. Berikut adalah *class diagram* prediksi diabetes yang dilihat pada Gambar



**Gambar 3. 13 Class Prediksi Diabetes**

### Perancangan Database

Rancangan basis data merupakan gambaran menyeluruh dari setiap tabel yang digunakan dalam pembangunan sistem. Rancangan ini juga menyertakan informasi atribut yang ada pada setiap field. Perancangan basis data ini dibuat dari kelas yang memerlukan penyimpanan data. Kelas yang memerlukan penyimpanan biasanya kelas yang memiliki atribut.

1. **Tabel Pasien**

Tabel dataset dimaksudkan untuk menyimpan data pasien yang akan digunakan untuk proses prediksi penyakit diabetes. Tabel dataset ditunjukan pada

**Tabel 3. 12 Tabel Pasien**

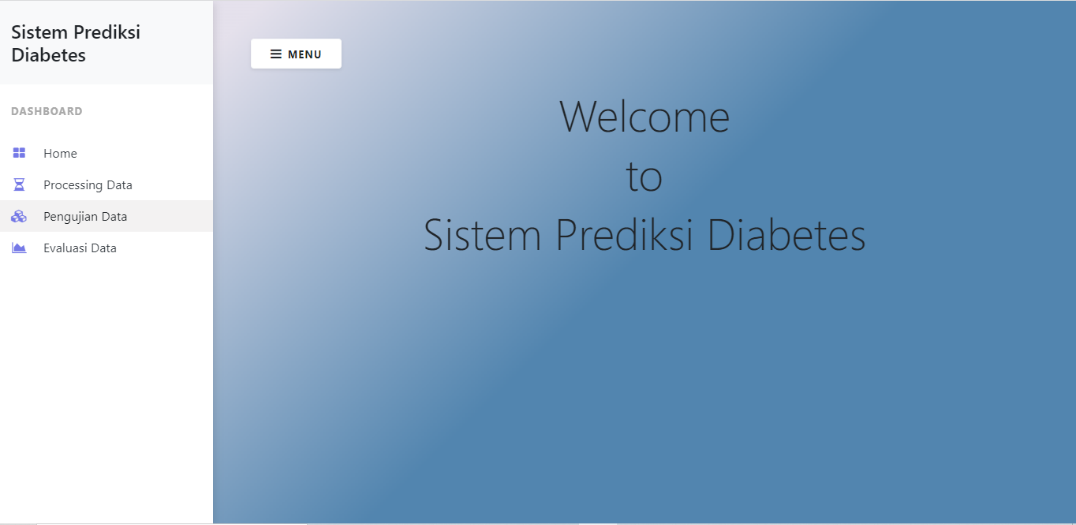
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No** | **Nama Field** | **Tipe** | **Ukuran** | **Keterangan** |
| 1 | id | Int | 11 | Primary Key |
| 2 | n\_pregnant | Float | 10 |  |
| 3 | glucose\_conc | Float | 10 |  |
| 4 | bp | Float | 10 |  |
| 5 | skin\_len | Float | 10 |  |
| 6 | insulin | Float | 10 |  |
| 7 | bmi | Float | 10 |  |
| 8 | pedigree\_fun | Float | 10 |  |
| 9 | age | Int | 11 |  |
| 10 | Output | Varchar | 11 |  |

### Perancangan Antarmuka

Perancangan antarmuka digambarkan dengan tampilan dari sistem yang dibangun, dan menjadi acuan dibuatnya sistem. Berikut adalah rancangan dari sistem prediksi penyakit diabetes.

1. **Antarmuka Home**

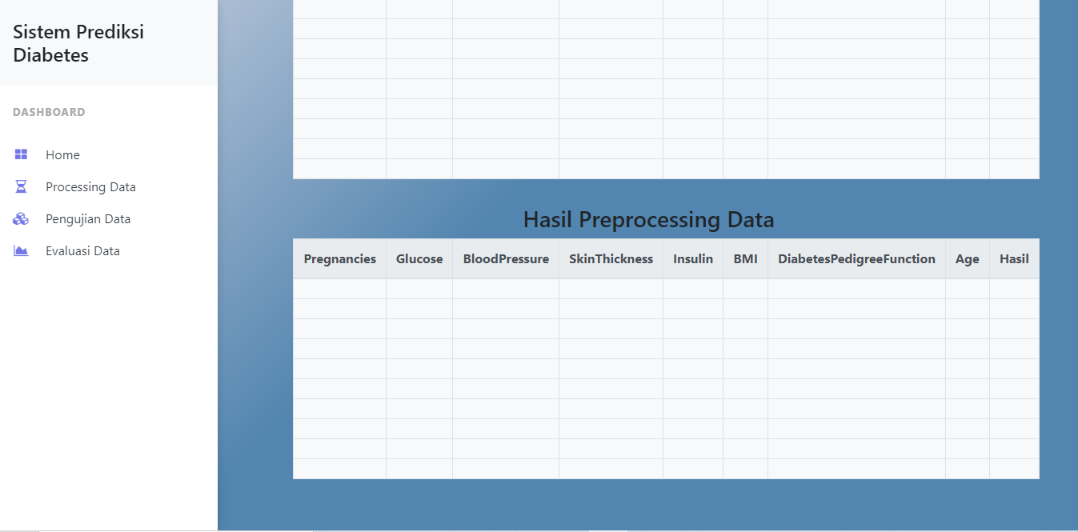
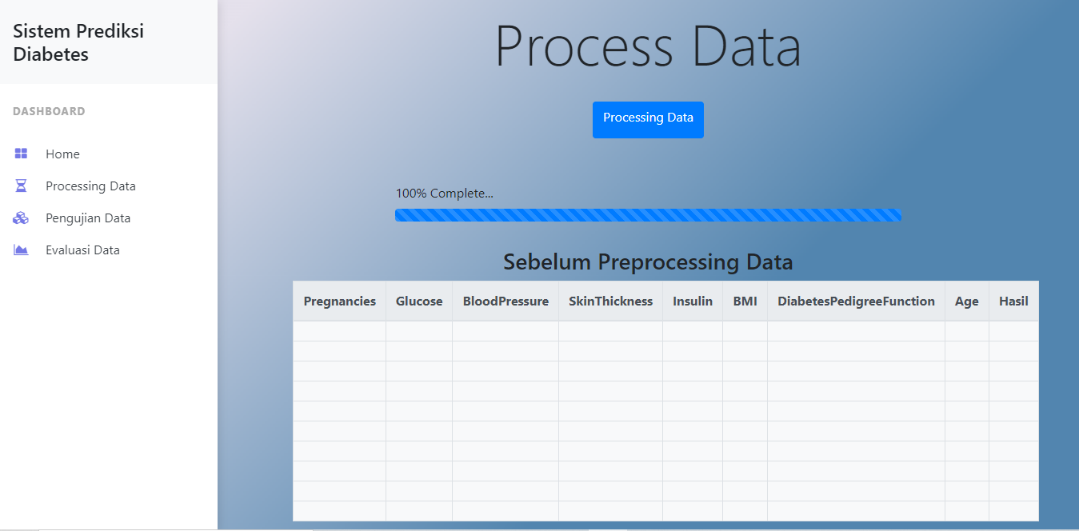
Halaman utama adalah tampilan awal saat program dijalankan. Halaman ini menampilkan 5 menu: Home, Processing Data, Pengujian Data, Evaluasi Data, serta Tambah Data. Dapat dilihat pada Gambar 3.14 Antarmuka Home.



**Gambar 3. 14 Antarmuka Home**

1. **Antarmuka Processing Data**

Halaman *processing* data adalah halaman untuk memprocess dataset yang digunakan, kemudian menampilkan hasil dataset kedalam tabel. Dapat dilihat pada Gambar 3.15 Antarmuka Processing Data

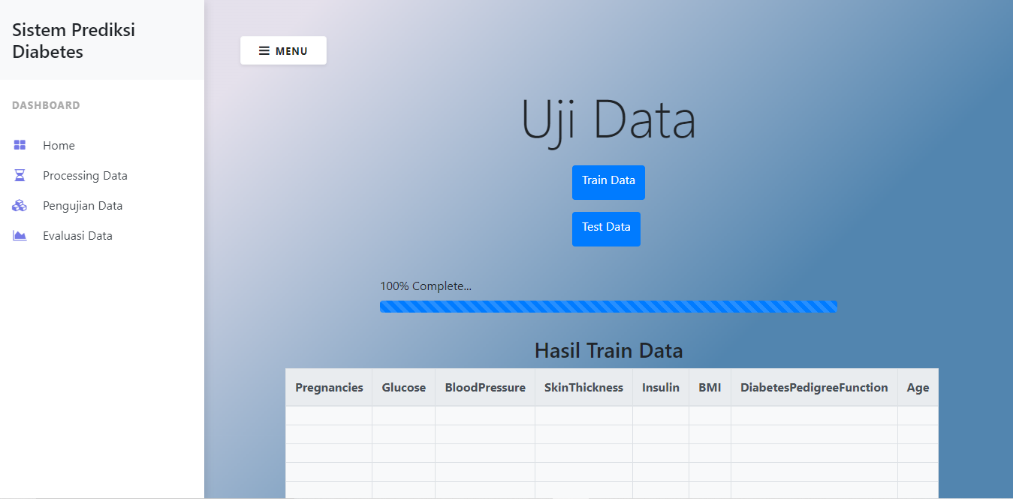


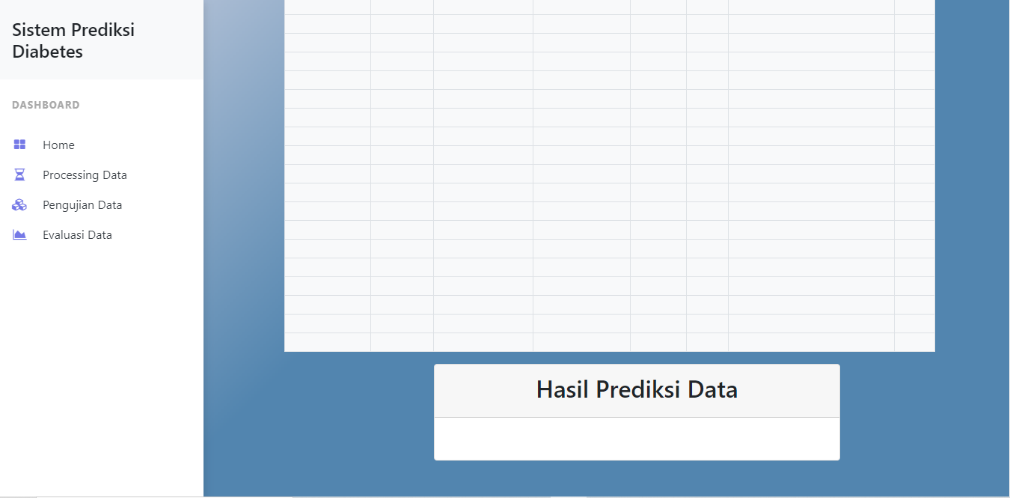
**Gambar 3. 15 Antarmuka Processing Data**

1. **Antarmuka Pengujian Data**

Halaman pengujian data merupakan halaman untuk memprediksi data yang telah diprocess pada halaman *processing*. Antarmuka pengujian dibagi menjadi 2 yaitu antarmuka data latih dan antarmuka data uji.

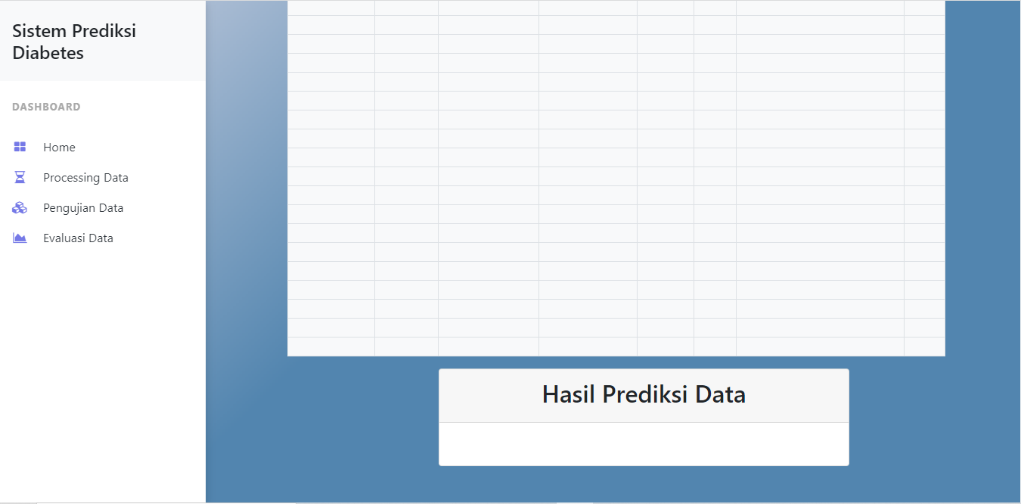
* **Data Latih**





**Gambar 3. 16 Antarmuka Data Latih**

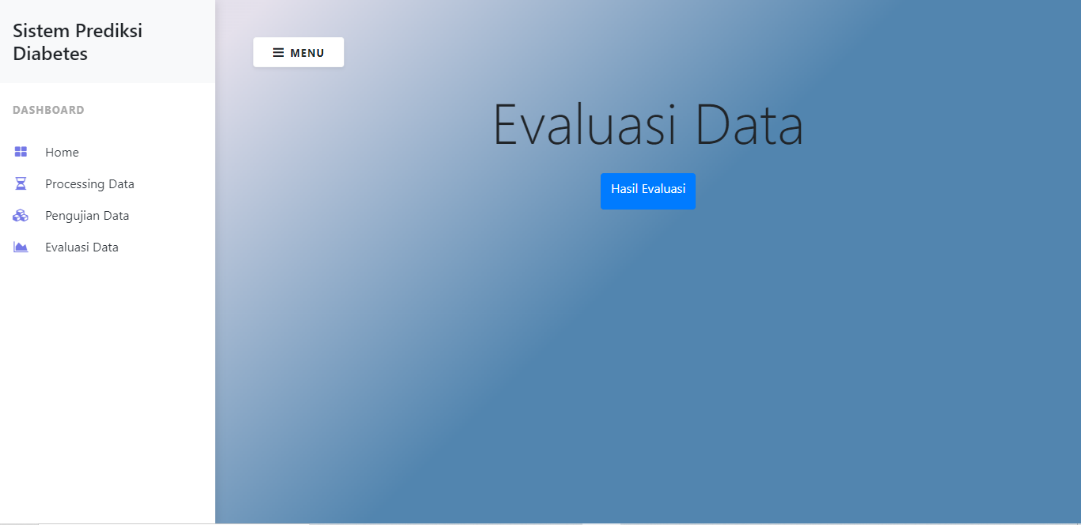
* **Data Uji**



**Gambar 3. 17 Antarmuka Data Uji**

1. **Antarmuka Evaluasi Data**

Pada rancangan ini, dimana setelah melakukan pengujian terhadap dataset yang digunakan, maka antarmuka ini diperuntukan menghitung hasil akurasi yang didapatkan dari model prediksi. Dapat dilihat pada Gambar 3.18 Antarmuka Evaluasi Data

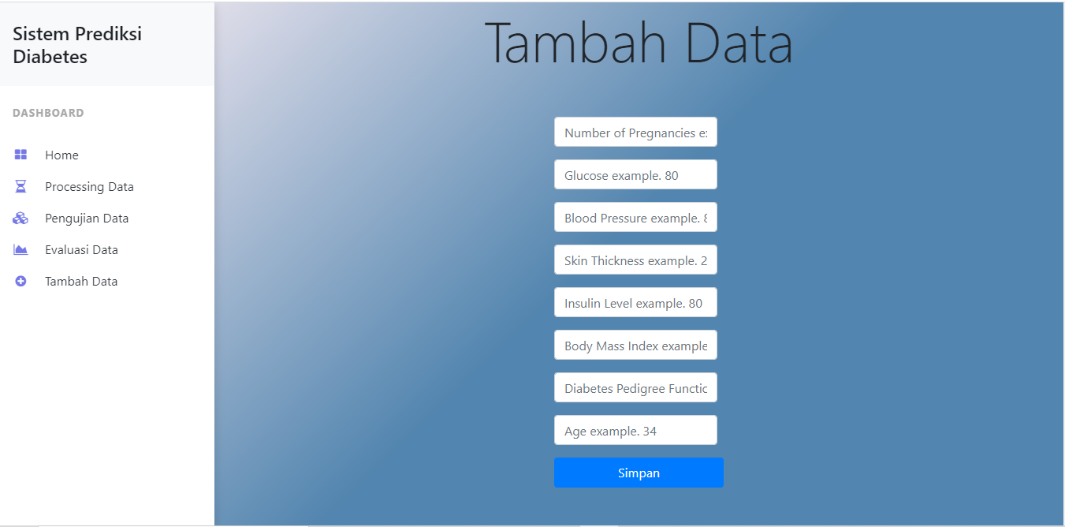




**Gambar 3. 18 Antarmuka Evaluasi Data**

1. **Antarmuka Tambah Data**

Pada rancangan ini, dimana *user* menginput data/menambahkan data baru ke dalam database. Dapat dilihat pada Gambar 3.19 Antarmuka Tambah Data



**Gambar 3. 19 Antarmuka Tambah Data**

# BAB IV

# IMPLEMENTASI DAN PENGUJIAN

## Implementasi Sistem

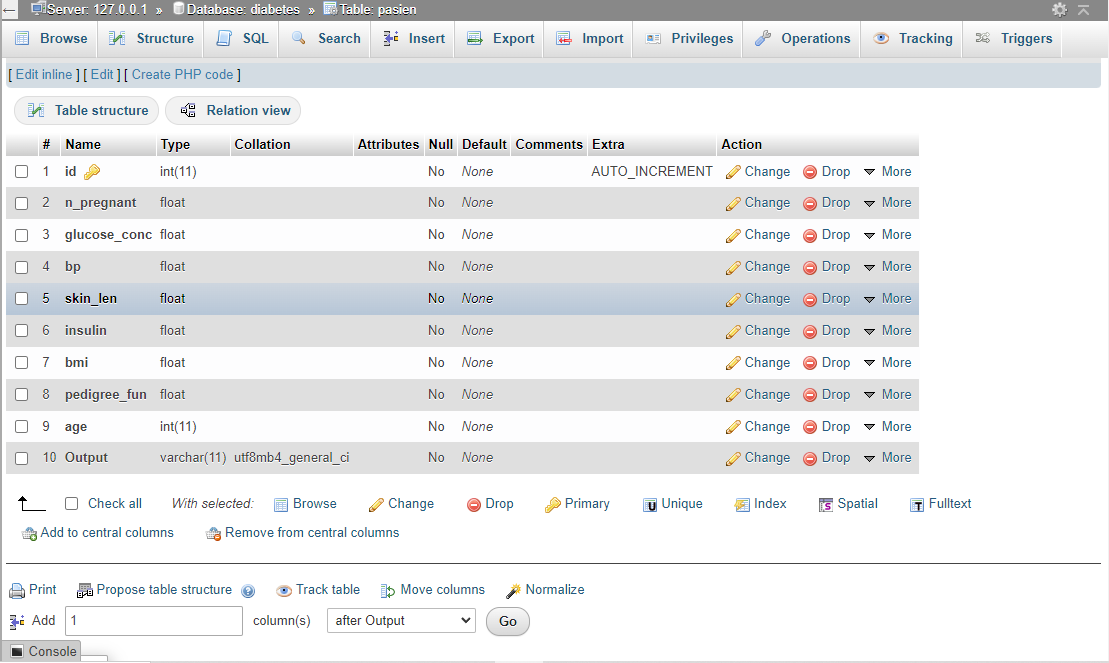
Pada penelitian ini, implementasi dan pengujian dilakukan dengan tanpa adanya koneksi internet. Hal tersebut dilakukan agar dapat kemudahan dalam melakukan pengelolaan dalam membangun sistem dan uji kualitas sistem. Pembangunan sistem prediksi penyakit diabetes ini dibangun dengan berbasis *dekstop* menggunakan bahasa pemograman *python* dan framework mvc *flask*, dan menggunakan *XAMPP* untuk koneksi *database MySQL*.

## Implementasi Database

Implementasi basis data yang dibangun berdasarkan perancangan basis data yang terdapat pada bab sebelumnya. Basis data ini dibuat dengan menggunakan aplikasi *MySQL*. Berikut merupakan penjelasan mengenai implementasi dari tiap tabel yang terdapat pada sistem prediksi penyakit diabetes.

* Database Pasien

Hasil perancangan tabel pasien pada bab sebelumnya di implementasikan pada *MySQL*. Implementasi *database* dapat dilihat pada Gambar 4.1 Database Pasien

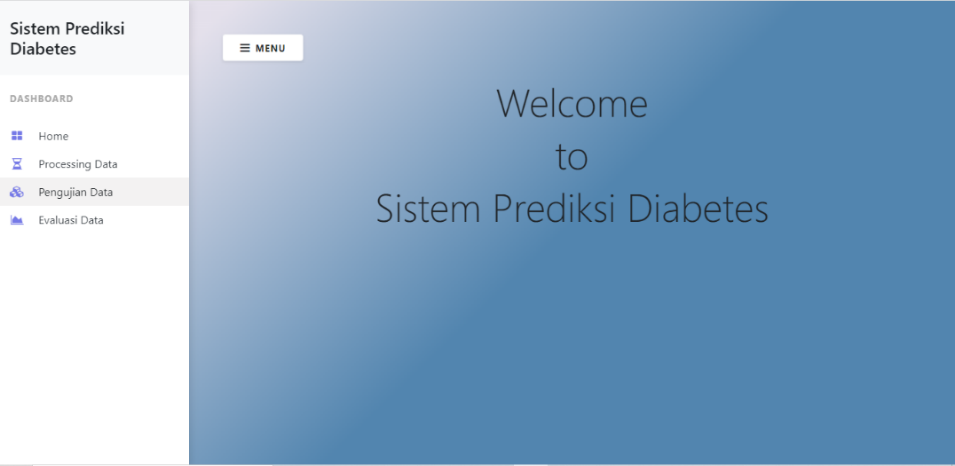


**Gambar 4. 1 Database Pasien Diabetes**

## Implementasi Antarmuka

1. Halaman Antarmuka Home

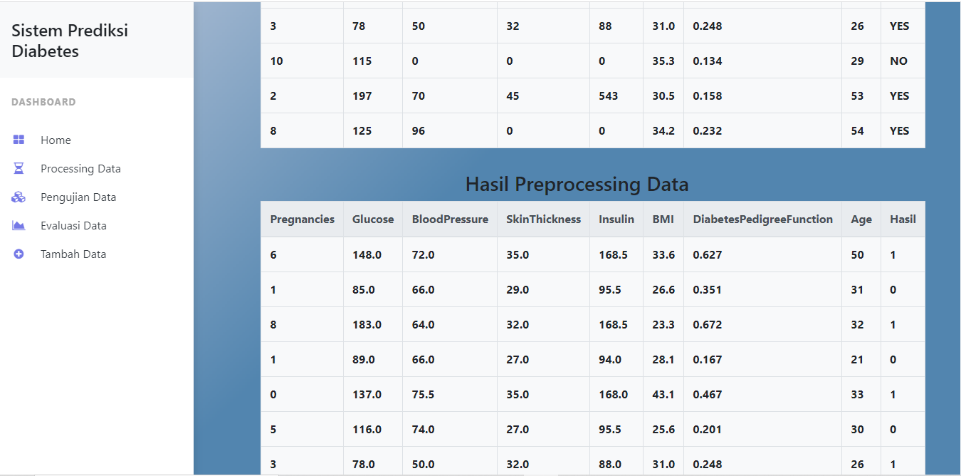
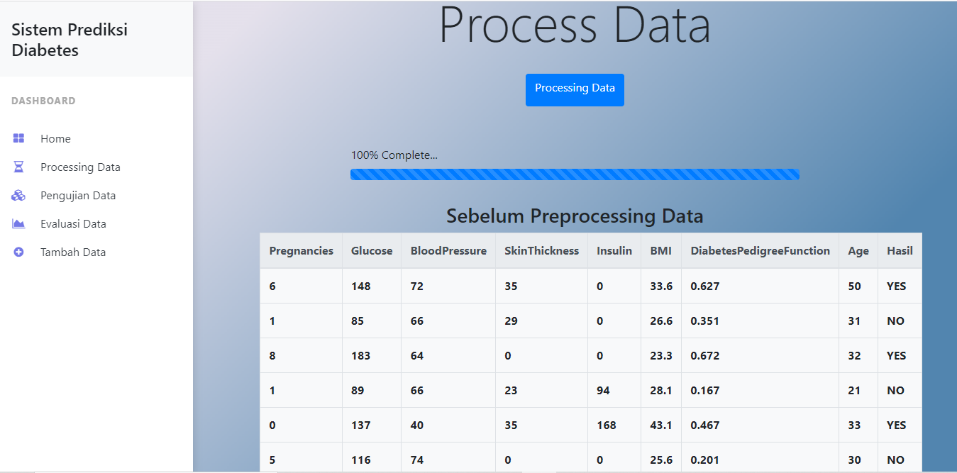
Antarmuka Home ini adalah tampilan awal saat sistem dijalankan yang dapat dilihat pada Gambar 4.2 Implementasi Antarmuka Home



**Gambar 4. 2 Implementasi Antarmuka Home**

1. Halaman Antarmuka Processing Data

Antarmuka Processing Data merupakan tampilan halaman dengan memprocess data yang digunakan dalam penelitian ini. Tiap atribut pada data diprocess dan ditampilkan dalam bentuk tabel. Dapat dilihat pada Gambar 4.3 Implementasi Antarmuka Processing Data

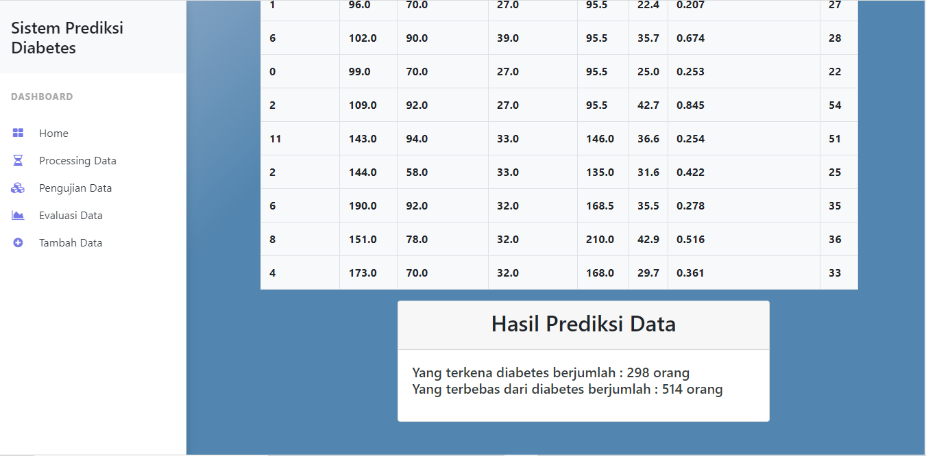
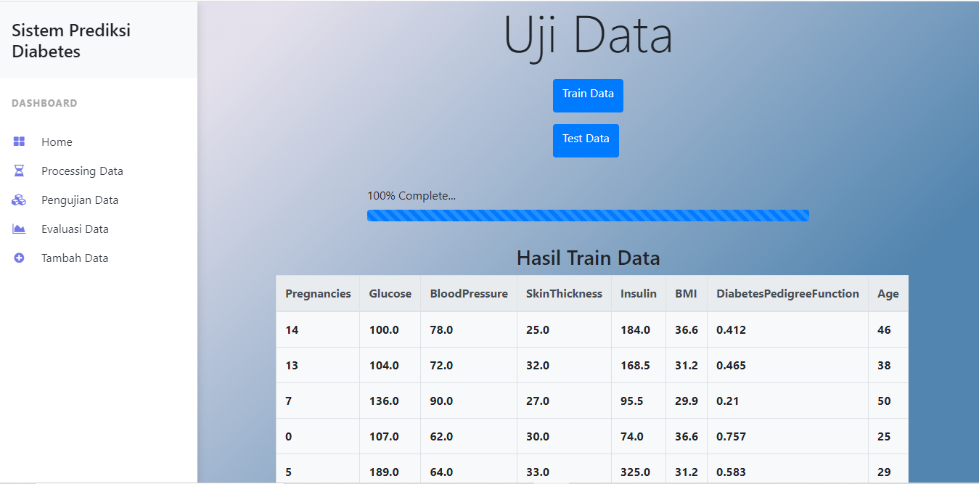


**Gambar 4. 3 Implementasi Antarmuka Processing Data**

1. Halaman Antarmuka Pengujian Data

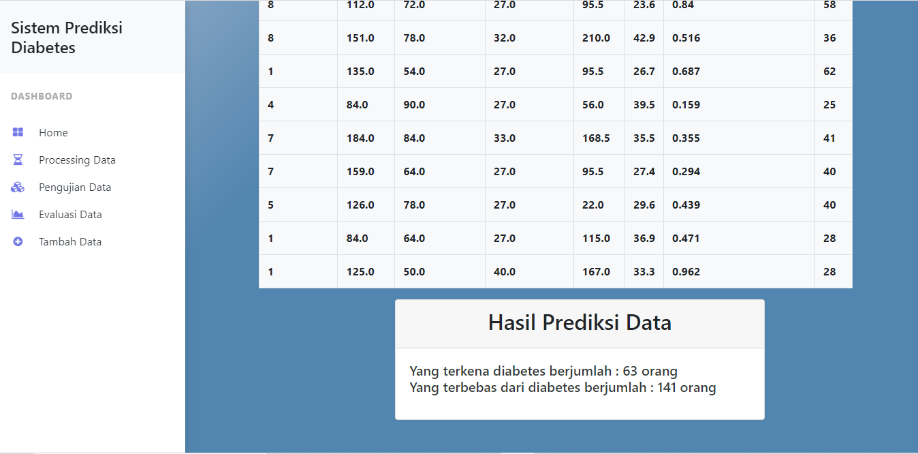
Antarmuka Pengujian Data merupakan tampilan halaman dimana, data yang telah melalui tahap processing akan diprediksi dan ditampilkan pada halaman pengujian data. Tampilan pada halaman ini dibagi menjadi 2 antarmuka yaitu halaman data latih dan halaman data uji.

* Data Latih



**Gambar 4. 4 Implementasi Antarmuka Data Latih**

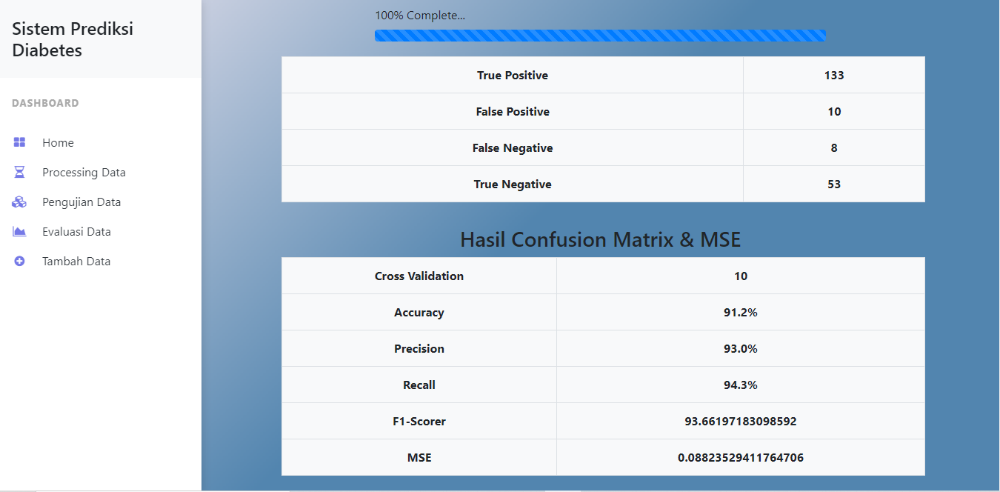
* Data Uji



**Gambar 4. 5 Implementasi Antarmuka Data Uji**

1. Halaman Antarmuka Evaluasi Data

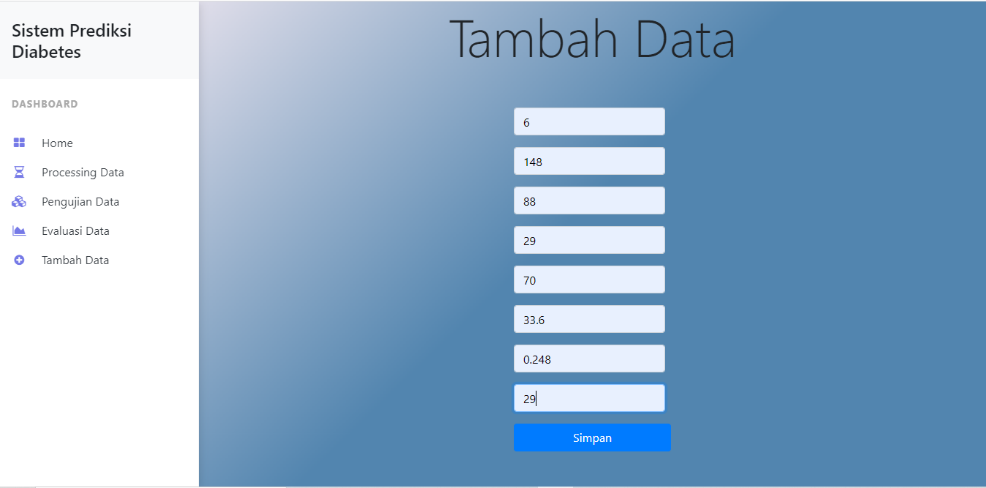
Antarmuka Evaluasi Data adalah tampilan halaman dimana hasil data yang telah diprediksi oleh model yang digunakan dalam penelitian ini, diuji untuk menentukan hasil akurasi dari model tersebut. Dapat dilihat pada Gambar 4.6 Implementasi Antarmuka Evaluasi Data



**Gambar 4. 6 Implementasi Antarmuka Evaluasi Data**

1. Halaman Antarmuka Tambah Data

Antarmuka Tambah Data merupakan tampilan halaman untuk menambahkan data baru ke dalam *database*. Berikut adalah Gambar 4.7 Implementasi Tambah Data



**Gambar 4. 7 Implementasi Antarmuka Tambah Data**

## Pengujian Akurasi

Pengujian akurasi merupakan tahapan untuk menguji hasil akurasi yang didapatkan pada model yang digunakan. Pengujian tersebut menggunakan metode *confusion matrix*, *mean square error*(mse).

1. Pengujian Confusion Matrix

Pengujian hasil prediksi yang dilakukan oleh model svm dan *forward selection* diuji menggunakan metode *confusion matrix*. Berikut adalah nilai *confusion matrix* yang dapat dilihat dibawah ini.

**Tabel 4. 1 Tabel Nilai Confusion Matrix**

|  |  |  |  |
| --- | --- | --- | --- |
| **True Positive (TP)** | **False Positive (FP)** | **False Negative (FN)** | **True Negative (TN)** |
| 133 | 10 | 8 | 53 |

Rumus:

Hasil:

1. Pengujian MSE

Pengujian *mean square error* dilakukan dengan menguji data yang digunakan dalam penelitian ini, dimana dalam kurung pangkat 2 terdapat At sebagai nilai aktual permintaan dikurangi dengan Ft sebagai nilai hasil peramalan. Kemudian dibagi dengan banyaknya data. Berikut adalah hasil pengujian.

Dari rumus pengujian dengan MSE mendapat hasil 0.088 berdasarkan data yang digunakan dalam penelitian ini. Dari hasil tersebut bahwa hasil perhitungan MSE sudah cukup baik.

# BAB V

# KESIMPULAN DAN SARAN

## Kesimpulan

Berdasarkan pada penelitian ini dibuat menggunakan algoritma *support vector machine* dan *forward selection*. Dalam penelitian ini 80% dari data digunakan sebagai latihan dan 20% dari data digunakan sebagai data uji dan hasil akurasi yang diperoleh cukup tinggi dalam memprediksi penyakit diabetes berdasarkan atribut yang ada pada data. Kesimpulan dari penelitian ini adalah metode *forward selection* dan *support vector machine* merupakan metode prediksi yang cukup baik untuk diterapkan pada bidang kesehatan terutama dalam memprediksi penyakit diabetes yang mana penyakit tersebut dikategorikan sulit untuk dideteksi.

## Saran

Saran penelitian selanjutnya dikemudian hari penelitian ini dapat di implementasikan ke berbagai bidang seperti memprediksi lokasi, nilai akademik, serta riwayat kesehatan keluarga dapat diperhitungkan agar hasil akurasi yang didapat lebih akurat.

# DAFTAR ISI

|  |  |  |
| --- | --- | --- |
| [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] |  | S. Pangribowo, “Tetap Produktif, Cegah, dan Atasi Diabetes Melitus,” *InfoDatin,* 18 Febuari 2020. |
| [4] |  | 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. |
| [5] |  | 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. |
| [6] |  | 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. |
| [7] |  | 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. |
| [8] |  | 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. |
| [9] |  | 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. |
| [10] |  | 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. |
| [11] |  | 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. |
| [12] |  | D. S. S. Deepti Sisodia, “Prediction of Diabetes using Classification Algorithms,” *ELSEVIER,* vol. 13, no. 2, pp. 1578-1585, 2018. |
| [13] |  | A. Kowalczyk, SUPPORT VECTOR MACHINES, Morrisville: Syncfusion, 2017. |
| [14] |  | 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. |
| [15] |  | 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. |
| [16] |  | 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. |
| [17] |  | 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. |
| [18] |  | H. Harafani, “Forward Selection pada Support Vector Machine untuk Memprediksi Kanker Payudara,” *Jurnal Infortech,* vol. 1, no. 2, pp. 131-139, 2019. |
| [19] |  | S. S. R. A. P. Fatkhuroji, “PREDIKSI HARGA KEDELAI LOKAL DAN KEDELAI IMPOR DENGAN METODE SUPPORT VECTOR MACHINE BERBASIS FORWARD SELECTION,” *Jurnal Teknologi Informasi,* vol. 15, no. 1, pp. 61-77, 2019. |
| [20] |  | 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. |
| [21] |  | 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. |
| [22] |  | 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. |
| [23] |  | E. B. N. B. Damir Imamovic, “Prediction of mortality in patients with cardiovascular disease using data mining methods,” *IEEEXplore,* vol. 25, no. 2, pp. 18-20, 2020. |
| [24] |  | J. C. F. Joanne B. Cole, “Genetics of diabetes mellitus and diabetes complications,” *Nephrology,* vol. 16, no. 2, pp. 377-390, 2020. |
| [25] |  | M. S. Petrov, “Panorama of mediators in postpancreatitis diabetes mellitus,” *Gastroenterology,* vol. 36, no. 5, pp. 443-451, 2020. |
| [26] |  | Z. S. L. A. A. M. A. A. C. T. R. &. P. H. A. Maleesa M. Pathirana, “Association between metabolic syndrome and gestational diabetes mellitus in women and their children: a systematic review and meta-analysis,” *Springer,* vol. 71, pp. 310-320, 2020. |
| [27] |  | R. C. V. Anuja Kumari, “Classification Of Diabetes Disease Using Support Vector Machine,” *International Journal of Engineering Research and Applications,* vol. 3, no. 2, pp. 1797-1801, 2013. |
| [28] |  | I. W. N. A. G. F. K. Liaqat Ali, “LDA–GA–SVM: improved hepatocellular carcinoma prediction through dimensionality reduction and genetically optimized support vector machine,” *Springer,* vol. 33, no. 4, p. 2783–2792, 2020. |
| [29] |  | U. E. Elin Nurlia, “PENERAPAN FITUR SELEKSI FORWARD SELECTION UNTUK MENENTUKAN KEMATIAN AKIBAT GAGAL JANTUNG MENGGUNAKAN ALGORITMA C4.5,” *JUTIM (Jurnal Teknik Informatika Musirawas),* vol. 6, no. 1, pp. 42-50, 2021. |
| [30] |  | S. b. J. Pradeep Kandhasamy, “Performance Analysis of Classifier Models to Predict Diabetes Mellitus,” *ELSEVIER,* vol. 47, pp. 45-51, 2015. |
| [31] |  | J. J. Pangaribuan, “MENDIAGNOSIS PENYAKIT DIABETES MELITUS DENGAN MENGGUNAKAN METODE EXTREME LEARNING MACHINE,” *Jurnal ISD,* vol. 2, no. 2, pp. 32-40, 2016. |
| [32] |  | G. P. P. G. F. P. G. M. M. D. R. P. C. G. M. A. Frattola, “Time and frequency domain estimates of spontaneous baroreflex sensitivity provide early detection of autonomic dysfunction in diabetes mellitus,” *Diabetologia,* vol. 40, pp. 1470-1475, 1997. |
| [33] |  | S. S. Shara Kurnia Trisnawati, “Faktor Risiko Kejadian Diabetes Melitus Di Puskesmas Kecamatan Cengkareng Jakarta Barat Tahun 2012,” *Jurnal Ilmiah Kesehatan,* vol. 5, no. 1, pp. 5-11, 2013. |
| [34] |  | Y. T. N. F. F. Z. Fitrah Reynaldi, “Penyuluhan Pentingnya Pencegahan Penyakit Diabetes Sejak Dini Kepada Guru Perempuan TK Yaa Bunaaya I Gampong Ujong Drien,” *JPAI (Jurnal Perempuan dan Anak Indonesia) ,* vol. 2, no. 2, pp. 26-30, 2020. |
| [35] |  | P. K. J. Priyanka Sonar, “DIABETES PREDICTION USING DIFFERENT MACHINE LEARNING APPROACHES,” *IEEE Xplore (3rd International Conference on Computing Methodologies and Communication (ICCMC)),* vol. 25, no. 6, pp. 367-371, 2019. |
| [36] |  | C. C. M. E. S. Ahmet Kadir Arslan, “Different medical data mining approaches basedprediction of ischemic stroke,” *ELSEVIER,* vol. 130, pp. 87-92, 2016. |