

**KLASIFIKASI PENYAKIT DIABETES PADA PEREMPUAN
MENGUNAKAN METODE *SUPPORT VECTOR MACHINE* (SVM)**

(Skripsi)

Oleh

**Aulia Zahro Kholili
NPM 1967031001**



**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2023**

ABSTRACT

CLASSIFICATION OF DIABETES IN WOMEN USING THE SUPPORT VECTOR MACHINE (SVM) METHOD

By

Aulia Zahro Kholili

Diabetes is a common chronic disease caused by high blood sugar levels that exceed normal limits. In diabetics, the pancreas cannot produce the required amount of insulin. Diabetes is a disease that causes the highest death rate in Indonesia. Diabetes is potentially higher in women. This study analyses the classification of those with the potential to develop diabetes, especially women, based on available data. The data used is a dataset from the article The National Institute of Diabetes and Digestive and Kidney Diseases. There are many ways to diagnose diabetes, one of which is using the Support Vector Machine algorithm. This study uses four kernels in modelling using SVM, namely linear, polynomial, gaussian, and sigmoid, as well as several training and data testing divisions. The results obtained show the highest model accuracy rate of 87% when using a linear kernel with training and testing data divisions of 95% and 5%.

Keywords: Diabetes, Classification, Support Vector Machine.

ABSTRAK

KLASIFIKASI PENYAKIT DIABETES PADA PEREMPUAN MENGUNAKAN METODE *SUPPORT VECTOR MACHINE* (SVM)

Oleh

Aulia Zahro Kholili

Diabetes adalah penyakit kronis umum yang diakibatkan kadar gula darah yang tinggi dan melewati batas normal. Pada penderita diabetes, pankreas tidak mampu menghasilkan jumlah insulin yang diperlukan. Diabetes merupakan salah satu penyakit yang mengakibatkan angka kematian tertinggi di Indonesia. Diabetes berpotensi lebih tinggi pada perempuan. Untuk membantu mendeteksi penyakit diabetes lebih dini, penelitian ini mengkaji analisis mengenai klasifikasi siapa saja yang berpotensi terkena penyakit diabetes, khususnya pada perempuan, berdasarkan data yang tersedia. Data yang digunakan adalah dataset dari artikel *The National Institute of Diabetes and Digestive and Kidney Disease*. Terdapat banyak cara dalam mendiagnosa penyakit diabetes, salah satunya adalah menggunakan algoritma *Support Vector Machine*. Penelitian ini menggunakan empat kernel dalam melakukan pemodelan menggunakan SVM, yaitu linear, polinomial, gaussian, dan sigmoid serta beberapa pembagian *training* dan *testing* data. Hasil yang diperoleh menunjukkan tingkat akurasi model tertinggi sebesar 87% ketika menggunakan kernel linear dengan pembagian *training* dan *testing* data sebesar 95% dan 5%

Kata Kunci: Diabetes, Klasifikasi, *Support Vector Machine*.

**KLASIFIKASI PENYAKIT DIABETES PADA PEREMPUAN
MENGUNAKAN METODE *SUPPORT VECTOR MACHINE* (SVM)**

Oleh

Aulia Zahro Kholili

(Skripsi)

**Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA MATEMATIKA**

Pada

**Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2023**

Judul Skripsi

: **KLASIFIKASI PENYAKIT DIABETES
PADA PEREMPUAN MENGGUNAKAN
METODE SUPPORT VECTOR MACHINE
(SVM)**

Nama Mahasiswa

: **Aulia Zahro Kholifi**

Nomor Pokok Mahasiswa

: **1967031001**

Jurusan

: **Matematika**

Fakultas

: **Matematika dan Ilmu Pengetahuan Alam**

MENYETUJUI

1. Komisi Pembimbing



Dr. Ahmad Faisol, S.Si., M.Sc.
NIP 1980020 6200312 1003



Dr. Aang Nuryaman, S.Si., M.Si.
NIP 19740316 200501 1 001

2. Ketua Jurusan Matematika



Dr. Aang Nuryaman, S.Si., M.Si.
NIP 19740316 200501 1 001

MENGESAHKAN

1. Tim Penguji

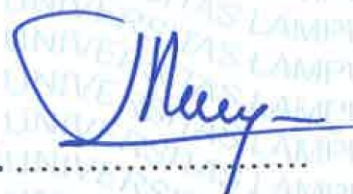
Ketua

: **Dr. Ahmad Faisol, S.Si., M.Sc.**



Sekretaris

: **Dr. Aang Nuryaman, S.Si., M.Si.**



Penguji

: **Drs. Eri Setiawan, M.Si.**



Bukan Pembimbing

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



Dr. Eng. Supto Dwi Yuwono, S.Si., M.T.

NIP 19740705 200003 1 001

Tanggal Lulus Ujian Skripsi : 8 Februari 2023

PERNYATAAN SKRIPSI MAHASISWA

Yang bertanda tangan di bawah ini :

Nama : **Aulia Zahro Kholili**
Nomor Pokok Mahasiswa : **1967031001**
Jurusan : **Matematika**
Judul Skripsi : **KLASIFIKASI PENYAKIT DIABETES
PADA PEREMPUAN MENGGUNAKAN
METODE *SUPPORT VECTOR*
MACHINE (SVM)**

Dengan ini menyatakan bahwa skripsi ini adalah hasil pekerjaan saya sendiri. Apabila kemudian hari terbukti bahwa skripsi ini merupakan hasil salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku.

Bandar Lampung, 8 Februari 2023

Yang menyatakan,



20
METERAI
TEMREL
C32AKX310058328

Aulia Zahro Kholili

NPM. 1967031001

Riwayat Hidup

Penulis bernama Aulia Zahro Kholili dan biasa disapa Aulia lahir di Banyuwangi, pada tanggal 21 Mei 2001, merupakan anak pertama dari empat bersaudara pasangan Bapak Mukholil dan Ibu Nur Isnainiyah.

Penulis menempuh pendidikan Sekolah Dasar di MIN Padang Ratu pada tahun 2007-2013. Kemudian melanjutkan pendidikan Sekolah Menengah Pertama di MTsN 3 Lampung Utara pada tahun 2013-2016 dan melanjutkan pendidikan Sekolah Menengah Atas di MAN 1 Metro pada tahun 2016-2019. Pada tahun 2019, penulis terdaftar sebagai mahasiswa S1 Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) Universitas Lampung melalui jalur Prestasi Khusus.

Selama menjadi mahasiswa, penulis aktif dalam organisasi Himpunan Mahasiswa Matematika (HIMATIKA). Selain aktif dalam organisasi kampus, pada tahun 2021 penulis mengikuti kegiatan Kredensial Mikro Mahasiswa Indonesia (KMMI) di Universitas Lampung, program Teknologi Multimedia. Pada tahun 2022 penulis mengikuti kegiatan Magang dan Studi Independen Bersertifikat (MSIB) yaitu Studi Independen Data Analyst pada bulan Agustus-Desember di PT. Mitra Talenta. Penulis juga melaksanakan Kuliah Kerja Nyata (KKN) di Desa Trimodadi, Kabupaten Lampung Utara, Lampung pada bulan Januari-februari 2022. Dan melaksanakan Kerja Praktik (KP) di Dinas Perpustakaan dan Kearsipan Kota Bandar Lampung pada bulan Juni-Juli 2022.

KATA INSPIRASI

"Allah tidak membebani seseorang melainkan sesuai dengan kesanggupannya"
(Q.S Al-Baqarah: 286)

"Maka janganlah kamu berkecil hati dan berputus asa, karena kamu lebih unggul jika kamu beriman"
(Q.S Al-Imron: 139)

Setiap orang memiliki proses untuk menjadi yang terbaik versi dirinya, jadi tidak perlu membandingkan prosesmu dengan orang lain

*Lakukanlah kebaikan sekecil apapun
karena kamu tidak pernah tahu kebaikan apa yang membawamu ke surge*
(Imam Hasan Al-Bashri)

Happiness is not by a chance, but by choice

*Everyone wants happiness, no one pain
But you cant have a rainbow without a little rain*

PERSEMBAHAN

Alhamdulillah, puji syukur kepada Allah SWT. yang telah memberikan petunjuk dan kekuatan juga memberikan penerangan dalam ilmu pengetahuan. Hanya karena-Nya lah skripsi ini bisa penulis selesaikan dengan rasa syukur dan bahagia. Dengan segala kerendahan hati, penulis persembahkan karya sederhana ini kepada:

Orang Tua Tercinta

Sebagai tanda terima kasih karena selalu mencurahkan doa, tenaga, pikiran, dan dukungannya untuk keberhasilan penulis dalam menuntut ilmu serta menjadi penyemangat terbaik sehingga penulis bisa menyelesaikan skripsi ini.

Dosen Pembimbing dan Pembahas

Yang senantiasa memberikan bimbingan, arahan, dan ilmu yang bermanfaat bagi penulis.

Almamaterku Tercinta, Universitas Lampung

SANWACANA

Puji syukur kehadirat Allah SWT yang telah memberikan limpahan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan tugas akhir berbentuk skripsi berjudul **“Klasifikasi Penyakit Diabetes pada Perempuan Menggunakan Metode *Support Vector Machine* (SVM)”**. Sholawat serta salam tak hentihentinya selalu tercurahkan kepada baginda besar kita Nabi Muhammad SAW yang kita nantikan syafaatnya di yaumul akhir kelak

Dalam penyusunan skripsi ini, penulis menyadari bahwa skripsi ini tidak akan terealisasi dengan baik tanpa adanya dukungan, bantuan, bimbingan serta arahan dari berbagai pihak. Pada kesempatan ini penulis ingin mengucapkan terima kasih sebesar-besarnya kepada:

1. Bapak Dr. Ahmad Faisol, S.Si., M.Sc., selaku dosen pembimbing I yang telah banyak memberikan kemudahan, nasihat, masukan dan saran, serta dorongan kepada penulis untuk menyelesaikan skripsi ini. Semoga ilmu yang bapak berikan selalu mendapat berkah dari Allah SWT dan bapak senantiasa dalam lindungan-Nya.
2. Bapak Dr. Aang Nuryaman, S.Si., M.Si., selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung dan Pembimbing II saya yang senantiasa membimbing dan memberikan arahan. Semoga bapak dimudahkan dalam segala urusan dan selalu dalam lindungan Allah SWT.
3. Bapak Drs. Eri Setiawan, M.Si., selaku dosen pembahas skripsi saya yang sudah memberikan kemudahan dalam menyelesaikan skripsi saya, dan

memberikan arahan serta nasihat untuk kebaikan hasil skripsi saya. Semoga bapak selalu dalam keadaan sehat dan selalu dalam lindungan Allah SWT.

4. Prof. Dra. Wamiliana, M.A., P.h.d., selaku dosen pembimbing akademik saya, yang telah membimbing selama masa perkuliahan.
5. Bapak Dr. Eng. Suropto Dwi Yuwono, S.Si., M.T., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
6. Seluruh Dosen, staf dan karyawan jurusan Matematika FMIPA Universitas Lampung.
7. Teristimewa kedua orangtuaku, Ayah Mukholil dan Ibu Nur Isnainiyah yang tak henti-hentinya mencurahkan doa, ridho, kasih sayang dan cinta untuk saya hingga sampai saat ini. Membimbing, mendidik dan menjadi panutan untuk saya mencapai segala mimpi serta cita-cita. Memberikan motivasi dan semangat untuk bangkit saat apa yang terjadi tidak sesuai dengan ekspektasi. Rela berkorban dan berjuang untuk sampai pada tahap ini, sungguh semua yang telah kalian beri untuk tiada terbalas. Semoga Allah selalu menjaga, melindungi, meridhoi, dan mengumpulkan kita beserta keluarga di syurganya kelak.
8. Adik-adikku tersayang Afif Adila Kholili, Muthia Sabrina Kholili, dan Kamila Nur Fadia Kholili terimakasih telah menjadi adik dan teman bermain yang asik. Semoga selalu dalam lindungan Allah SWT, dimudahkan dalam segala urusan, dapat bermanfaat bagi banyak orang, dan membanggakan kedua orangtua serta keluarga.
9. Bapak Drajat yang selalu memberi dukungan dan bantuan dalam kelancaran seminar proposal, seminar hasil, hingga kompre. Terimakasih bapak, jasa bapak tidak akan terlupakan.
10. Rekan mahasiswa Matematika Universitas Lampung angkatan 2019.
11. Sahabat-sahabat keluarga Annisa, Andini, Karla, Renfilia, Yumna, Sabilla, Luluk, Natasya, Annisa, Lutfia, Nur Hidayah, dan Syifa yang selalu memberi semangat dalam penyelesaian skripsi.
12. Sahabat terbaikku Iin Nailiya Suherdi, yang selalu menjadi pendengar semua keluh kesah selama di pesma. Terimakasih telah menjadi sosok kakak, dan sahabat yang selalu ada dan bisa diandalkan, selalu memberi contoh dan

menjadi panutan yang baik, memberi motivasi dan nasihat serta dukungan. Terimakasih banyak untuk semua ilmu kehidupan yang telah diberikan.

13. Teman perskripsianku, Qorry Indira Devina, terimakasih telah menjadi partner pejuang skripsi dari awal hingga akhir. Semoga Allah SWT selalu memberi kelancaran untuk segala urusan kita.
14. Teman-teman seperjuangan, Listra, Eccha, Hana, Ulan, Shella, Meli, Nada, Alenia, Aulia Ayu, Deswita, Hijri, Aprici, Febby, Debi, Triya, Roro dan Upi atas bantuan, dukungan, serta saran dalam pengerjaan skripsi ini.
15. Keluarga besar Pesma Al- Huda Putri angkatan II, Iin Nailiya Suherdi (kak Iin), Irgi Cleareassa Anastasya Harend (kak Irgi), Lailatul Farihah (kak Tufa), Lita Putri (kak Lita), terimakasih selalu ada disaat senang dan susah, serta menjadi pengingat disaat lalai, suatu kebahagiaan dapat mengenal kalian, semoga Allah memudahkan segala urusan kalian dan mengumpulkan kita di jannah-Nya.
16. Semua pihak yang membantu penyelesaian skripsi ini yang tidak mungkin penulis menyebutkan satu per satu.

Akhir kata, semoga Allah SWT senantiasa melimpahkan Rahmat-Nya dan membalas segala kebaikan kepada pihak-pihak yang telah membantu penulis. Penulis menyadari bahwa laporan ini masih banyak kekurangan dalam menuliskan skripsi ini, namun penulis berharap dengan adanya laporan ini dapat bermanfaat bagi setiap pembaca.

Bandar Lampung, 8 Februari 2023

Penulis

Aulia Zahro Kholili

DAFTAR ISI

	Halaman
ABSTRAK	ii
DAFTAR GAMBAR.....	xv
DAFTAR TABEL.....	xvii
I. PENDAHULUAN	
1.1 Latar Belakang Masalah	1
1.1 Tujuan	3
1.2 Manfaat	3
II. TINJAUAN PUSTAKA	
2.1 Diabetes	4
2.2 Klasifikasi	5
2.3 <i>Support Vector Machine</i> (SVM).....	5
2.4 <i>Percentage Split</i>	11
2.5 <i>Confusion Matrix</i>	12
III. METODOLOGI PENELITIAN	
3.1. Waktu dan Tempat Penelitian.....	14
3.2. Data Penelitian	14
3.3. Metode Penelitian	15
IV. HASIL DAN PEMBAHASAN	
4.1. Statistik Deskriptif	17
4.2. Regresi Logistik	19
4.3. Analisis SVM.....	23

V. PENUTUP

5.1 Kesimpulan	38
5.2 Saran	38

DAFTAR PUSTAKA

LAMPIRAN

DAFTAR GAMBAR

Halaman

1. *Hyperplane* Terbaik Yang Memisahkan Kedua *Class* Negatif Dan Positif.....6
2. Representasi Hasil Confusion Matrix.....12
3. *Flowchart* Algoritma *Support Vector Machine* (SVM)16

DAFTAR TABEL

TABEL	Halaman
1. Data penderita diabetes pada perempuan.....	7
2. Hasil klasifikasi data uji	11
3. Statistik deskriptif data kontinu	17
4. Statistik deskriptif data kategorik	18
5. <i>Iteration History</i>	19
6. Uji Signifikansi Parameter	20
7. Model Summary	20
8. Uji Parsial	21
9. Uji Kesesuaian Model	22
10. Hasil Klasifikasi	22
11. <i>Confusion matrix</i> kernel linear dengan dataset 80% dan 20%.....	24
12. <i>Classification report</i> kernel linear dengan dataset 80% dan 20%	25
13. <i>Confusion matrix</i> kernel linear dengan dataset 90% dan 10%	25
14. <i>Classification report</i> kernel linear dengan dataset 90% dan 10%	26
15. <i>Confusion matrix</i> kernel linear dengan dataset 95% dan 5%	26
16. <i>Classification report</i> kernel linear dengan dataset 95% dan 5%	27
17. <i>Confusion matrix</i> kernel polinomial dengan dataset 80% dan 20%	27
18. <i>Classification report</i> kernel polinomial dengan dataset 80% dan 20%	28
19. <i>Confusion matrix</i> kernel polinomial dengan dataset 90% dan 10%	28
20. <i>Classification report</i> kernel polinomial dengan dataset 90% dan 10%	29
21. <i>Confusion matrix</i> kernel polinomial dengan dataset 95% dan 5%	29
22. <i>Classification report</i> kernel polinomial dengan dataset 95% dan 5%	30
23. <i>Confusion matrix</i> kernel Gaussian RBF dengan dataset 80% dan 20%.....	30
24. <i>Classification report</i> kernel Gaussian RBF dengan dataset 80% dan 20% ..	31

25. <i>Confusion matrix</i> kernel Gaussian RBF dengan dataset 90% dan 10%.....	31
26. <i>Classification report</i> kernel Gaussian RBF dengan dataset 90% dan 10% ...	32
27. <i>Confusion matrix</i> kernel Gaussian RBF dengan dataset 95% dan 5%	32
28. <i>Classification report</i> kernel Gaussian RBF dengan dataset 95% dan 5%	33
29. <i>Confusion matrix</i> kernel sigmoid dengan dataset 80% dan 20%	33
30. <i>Classification report</i> kernel sigmoid dengan dataset 80% dan 20%.....	34
31. <i>Confusion matrix</i> kernel sigmoid dengan dataset 90% dan 10%	34
32. <i>Classification report</i> kernel sigmoid dengan dataset 90% dan 10%.....	35
33. <i>Confusion matrix</i> kernel sigmoid dengan dataset 95% dan 5%.....	35
34. <i>Classification report</i> kernel sigmoid dengan dataset 95% dan 5%	36
35. Model terbaik dengan metode SVM	36

I. PENDAHULUAN

1.1 Latar Belakang

Diabetes adalah penyakit kronis umum yang menyebabkan kadar gula darah tinggi. Segala usia dipengaruhi oleh penyakit ini, yang disebabkan oleh masalah mekanisme kontrol gula darah tubuh. Tubuh pasien diabetes tidak dapat menggunakan glukosa. Pankreas mengeluarkan hormon insulin, yang mengatur kadar gula darah. Pada penderita diabetes, pankreas tidak mampu menghasilkan jumlah insulin yang diperlukan. Tanpa insulin, sel-sel tubuh tidak dapat menyerap glukosa dan menggunakannya sebagai bahan bakar. Glukosa darah menumpuk ketika tidak diserap secara efektif oleh sel-sel tubuh. Organ tubuh dapat mengembangkan sejumlah penyakit sebagai akibat dari kondisi ini. Diabetes dapat menyebabkan masalah yang mengancam jiwa bagi pasien jika tidak dikelola secara memadai.

Dalam banyak kasus diabetes dapat dikontrol melalui diet dan olahraga teratur. Namun, kasus yang parah memerlukan intervensi medis, seperti suntikan insulin. Ada beberapa jenis diabetes, itulah sebabnya diabetes menjadi topik hangat di bidang medis. Diabetes memiliki dampak buruk pada tubuh dan pikiran. Kondisi ini membatasi kemampuan untuk berolahraga dan makan dengan sehat sehingga dapat menyebabkan obesitas dan penyakit terkait lainnya. Diabetes merupakan salah satu penyakit kronis yang menyebabkan angka kematian tertinggi di Indonesia, bahkan jumlah kasus nyerinya terus meningkat.

Sejak tahun 1980, rata-rata prevalensi tahunan diabetes pada orang di seluruh dunia telah meningkat dari 4,7% menjadi 8,5%. Pada tahun 2012, diabetes menyumbang 1,5 juta kematian. 43% dari 3,7 juta kematian terjadi sebelum usia 70 tahun, dan 2,2 juta dari kematian tersebut terkait dengan kadar glukosa darah yang berada di luar kisaran yang dapat diterima, yang juga meningkatkan risiko penyakit kardiovaskular dan gangguan lainnya. Oleh karena itu diperlukan suatu sistem untuk mengklasifikasikan data pasien untuk mengendalikan diabetes dengan mengidentifikasinya sejak dini.

Penelitian sebelumnya telah banyak dilakukan dengan metode klasifikasi contohnya pada penelitian Raudlatul (2016) yang melakukan penelitian Penerapan Metode *Support Vector Machine* pada Diagnosa Hepatitis. Penelitian ini bertujuan untuk mengetahui diagnosa penyakit hepatitis berdasarkan gejalanya menggunakan dua kernel untuk pengujian, yaitu kernel linier dengan tingkat akurasi 68-83% dan fungsi kernel *Radial Basic Function* (RBF) 70-96%.

Sementara itu, Dendry, et al., (2019) telah meliput kasus diabetes dengan metode *Neighbor Weighted K-Nearest Neighbor* pada Klasifikasi Diabetes Mellitus. Untuk membantu konsumen membedakan diabetes melitus tipe 1 dan 2, penelitian ini akan mengklasifikasikan diabetes melitus berdasarkan gejalanya. Dalam penelitiannya yang berjudul Implementasi *Support Vector Machine* (SVM) untuk Klasifikasi Penderita Diabetes Mellitus, Favorisen, et al., (2022) menyebutkan kasus lebih lanjut yang sebanding. Berdasarkan gejalanya, diabetes melitus akan dikategorikan dalam penelitian ini. Data yang digunakan berasal dari situs web *UCI Machine Learning Repository* dan dari *Diabetes 130-US Hospitals For Years 1999-2008 Dataset*. Dataset memiliki 101766 catatan dengan 34 atribut, yakni *Race, Gender, Age, Time In Hospital, Diag_1, Diag_2, Diag_3, Max_glu_serum, Metformin, Repaglinide, Nateglinide, Chlorpropamide, Glimepiride, Acetohexamide, Glipizide, Glyburide, Tolbutamide, Pioglitazone, Rosiglitazone*

Acarbose, Miglitol, Troglitazone, Tolazamide, Examide, Citoglipton, Insulin, Glyburidemetformin, Glipizide-metformin, Glimepiride-pioglitazone, Metformin-Rosiglitazone, Metforminpioglitazone, DiabetesMed, Change, Readmitted.

Penelitian-penelitian yang sudah dilakukan yaitu klasifikasi penyakit diabetes secara umum pada laki-laki dan perempuan. Oleh karena itu penulis tertarik untuk melakukan penelitian klasifikasi penyakit diabetes secara khusus pada perempuan, dengan harapan masyarakat perempuan dapat mengantisipasi diabetes lebih dini dengan mengenali gejalanya dan dapat membantu tenaga medis dalam diagnosa terhadap pasiennya.

1.2 Tujuan Penelitian

Adapun tujuan yang ingin dicapai dari penelitian ini adalah:

1. Penerapan metode *Support Vector Machine* (SVM) dalam klasifikasi penyakit diabetes pada perempuan.
2. Mendapatkan *hyperplane* terbaik dalam klasifikasi penyakit diabetes pada perempuan dengan metode *Support Vector Machine* (SVM).

1.3 Manfaat Penelitian

Manfaat yang diharapkan dalam penelitian ini adalah :

1. Memberikan informasi mengenai diabetes.
2. Menambah pengetahuan mengenai metode *Support Vector Machine*.
3. Membantu masyarakat perempuan untuk penanggulangan diabetes lebih dini.

II. TINJAUAN PUSTAKA

2.1 Diabetes

Diabetes atau yang sering dikenal dengan diabetes melitus merupakan suatu kondisi kronis yang ditandai dengan peningkatan glukosa darah atau kadar gula (Devi, et al., 2019). Diabetes adalah kondisi yang sangat berbahaya. Dampak jangka panjang diabetes muncul secara bertahap. Risiko komplikasi meningkat dengan durasi diabetes dan derajat disregulasi gula darah. Akhirnya, komplikasi dari diabetes bisa berakibat fatal atau bahkan melumpuhkan. Kemungkinan efek samping termasuk penyakit kardiovaskular, kerusakan saraf, kerusakan ginjal, kerusakan mata, kerusakan kulit, gangguan pendengaran, dan bahkan depresi (Clinic, 2018).

Diabetes melitus yang fatal adalah kondisi yang diturunkan dalam keluarga. Gaya hidup sehat tidak penting bagi sebagian besar individu. Aspek kehidupan yang paling krusial adalah kesehatan, namun hanya sedikit orang yang menyadari masalah kesehatan yang dapat meningkatkan risiko kematian (Fernanda, et al., 2017). Karena ketidaktahuan mereka tentang diabetes mellitus dan risiko yang terkait dengan manajemen yang tidak tepat, sebagian besar orang tidak mengetahui apakah mereka menderita diabetes tipe 1 atau tipe 2. Selain itu, sejumlah penyakit memiliki tanda-tanda awal yang mirip dengan diabetes melitus. Bagi penderita diabetes melitus, akan berakibat fatal jika salah menganalisisnya.

2.2 Klasifikasi

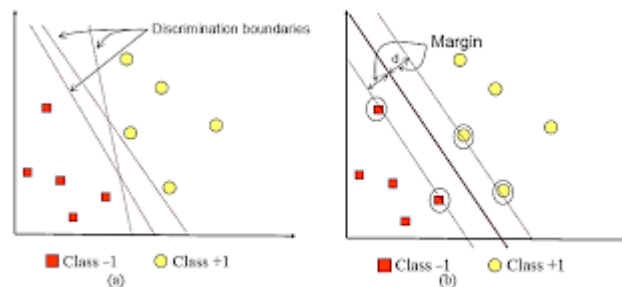
Untuk menggunakan model untuk meramalkan kelas suatu objek ketika label kelas tidak diketahui, klasifikasi adalah proses mengidentifikasi model atau fungsi yang menjelaskan dan membedakan kelas atau konsep dari data. Model saat ini dibuat menggunakan data latih dan data uji (Han dan Kamber, 2006). Algoritma *Naive Bayes*, *Adaptive Bayes Network*, *Decision Tree*, dan *Support Vector Machine* dapat digunakan untuk klasifikasi.

2.3 *Support Vector Machine* (SVM)

Vapnik pertama kali mempresentasikan *Support Vector Machine* (SVM) pada tahun 1992 sebagai teknik pembelajaran mesin berdasarkan gagasan *Structural Risk Minimization* (SRM), yang berupaya mengidentifikasi *hyperplane* optimal untuk membagi dua kelas pada ruang input. Metode ini menerapkan bias pembelajaran yang berasal dari teori pembelajaran statistik dan mengandalkan asumsi fungsi linier dalam ruang fitur berdimensi tinggi. Fungsi dan parameter kernel yang digunakan selama fase *switching* SVM memiliki dampak yang signifikan terhadap kebenaran model yang akan dihasilkan (Siagian, 2011).

Metode SVM diklasifikasikan menjadi dua kategori, yaitu SVM linier dan SVM nonlinier, berdasarkan sifat-sifatnya. Data yang dibagi secara linear, atau yang membagi dua kelas pada *hyperplane* dengan margin lunak, adalah yang dimaksud dengan SVM linier. SVM bersifat non-linier, khususnya ketika menerapkan fungsi kerneltrick pada ruang berdimensi tinggi (Rachman, 2012).

Teori SVM dapat dijelaskan secara gamblang sebagai strategi pencarian *hyperplane* terbaik yang berfungsi sebagai pemisah dua bagian kelas pada input *space*. Gambar 1a menampilkan beberapa piksel yang mewakili anggota dari dua kelas positif (dinotasikan +1) dan negatif (dinotasikan -1). Proses pembelajaran tentang klasifikasi masalah dipandang sebagai sarana untuk menemukan garis *hyperplane* yang menghubungkan dua kelompok tersebut. Sejumlah pemisah (dekriminalisasi batas) lainnya ditunjukkan pada Gambar 1. (Nugroho, 2008).



Gambar 1. *Hyperplane* Terbaik Yang Memisahkan Kedua *Class* Negatif Dan Positif

Upaya pencarian *hyperplane* yang optimal antara dua kelas pada ruang input adalah gagasan SVM. Pada gambar 1 menunjukkan pola milik dua kelas potongan: +1 dan -1. Lingkaran kuning dan kotak merah masing-masing mewakili pola yang terhubung di kelas -1 dan kelas +1. Dimungkinkan untuk memahami masalah dengan klasifikasi ini dengan mencoba mengidentifikasi garis-*hyperplane* yang membagi dua kelompok. *Hyperplane* terbaik, yang terletak di tengah hingga tengah kelas kedua, digambarkan pada gambar 1 dengan garis padat, sedangkan vektor pendukung diwakili oleh titik merah dan kuning di dalam lingkaran hitam.

Berikut merupakan persamaan untuk menentukan optimal *hyperplane* kedua kelas:

$$\text{Minimize } J_1[w] = \frac{1}{2} \|w\|^2 \quad (2.6)$$

$$\text{Dengan } y_i (x_i \cdot w + b) - 1 \geq 0, i = 1, \dots, n \quad (2.7)$$

Keterangan:

b = Nilai bias.

w = Bobot vektor.

x_i = Data ke-i.

y_i = Kelas data ke-i.

Diberikan contoh sebagai berikut:

Tabel 1. Data penderita diabetes pada perempuan

Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
6	148	72	35	0	33.6	0.627	50	1
1	85	66	29	0	26.6	0.351	31	-1
5	116	74	0	0	25.6	0.201	30	-1
3	78	50	32	88	31	0.248	26	1
10	115	0	0	0	35.3	0.134	29	-1
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	-1
10	168	74	0	0	38	0.537	34	1

Karena terdapat terdapat 8 variabel bebas yaitu *pregnancies*, *glucose*, *bloodpressure*, *skinthickness*, *insulin*, *BMI*, *diabetes pedigree function* dan *age* maka akan dicari 8 bobot yaitu w_1 , w_2 , w_3 , w_4 , w_5 , w_6 , w_7 dan w_8 . Sehingga akan diperoleh persamaan sebagai berikut:

$$a. \quad 1(6w_1 + 148w_2 + 72w_3 + 35w_4 + 0w_5 + 33.6w_6 + 0.627w_7 + 50w_8 + b) \geq 1$$

- b. $-1(1w_1 + 85w_2 + 66w_3 + 29w_4 + 0w_5 + 26.6w_6 + 0.351w_7 + 31w_8 + b) \geq 1$
- c. $-1(5w_1 + 116w_2 + 74w_3 + 0w_4 + 0w_5 + 25.6w_6 + 0.201w_7 + 30w_8 + b) \geq 1$
- d. $1(3w_1 + 78w_2 + 50w_3 + 32w_4 + 88w_5 + 31w_6 + 0.248w_7 + 26w_8 + b) \geq 1$
- e. $-1(10w_1 + 115w_2 + 0w_3 + 0w_4 + 0w_5 + 35.3w_6 + 0.134w_7 + 29w_8 + b) \geq 1$
- f. $1(2w_1 + 197w_2 + 70w_3 + 45w_4 + 543w_5 + 30.5w_6 + 0.158w_7 + 53w_8 + b) \geq 1$
- g. $1(8w_1 + 125w_2 + 96w_3 + 0w_4 + 0w_5 + 0w_6 + 0.232w_7 + 54w_8 + b) \geq 1$
- h. $-1(4w_1 + 110w_2 + 92w_3 + 0w_4 + 0w_5 + 37.6w_6 + 0.191w_7 + 30w_8 + b) \geq 1$
- i. $1(10w_1 + 168w_2 + 74w_3 + 0w_4 + 0w_5 + 38w_6 + 0.537w_7 + 34w_8 + b) \geq 1$

Kemudian persamaan diatas diubah menjadi

- a. $(6w_1 + 148w_2 + 72w_3 + 35w_4 + 0w_5 + 33.6w_6 + 0.627w_7 + 50w_8 + b) \geq 1$
- b. $(-1w_1 - 85w_2 - 66w_3 - 29w_4 - 0w_5 - 26.6w_6 - 0.351w_7 - 31w_8 - b) \geq 1$
- c. $(-5w_1 - 116w_2 - 74w_3 - 0w_4 - 0w_5 - 25.6w_6 - 0.201w_7 - 30w_8 - b) \geq 1$
- d. $(3w_1 + 78w_2 + 50w_3 + 32w_4 + 88w_5 + 31w_6 + 0.248w_7 + 26w_8 + b) \geq 1$
- e. $(-10w_1 - 115w_2 - 0w_3 - 0w_4 - 0w_5 - 35.3w_6 - 0.134w_7 - 29w_8 - b) \geq 1$
- f. $(2w_1 + 197w_2 + 70w_3 + 45w_4 + 543w_5 + 30.5w_6 + 0.158w_7 + 53w_8 + b) \geq 1$
- g. $(8w_1 + 125w_2 + 96w_3 + 0w_4 + 0w_5 + 0w_6 + 0.232w_7 + 54w_8 + b) \geq 1$
- h. $(-4w_1 - 110w_2 - 92w_3 - 0w_4 - 0w_5 - 37.6w_6 - 0.191w_7 - 30w_8 - b) \geq 1$
- i. $(10w_1 + 168w_2 + 74w_3 + 0w_4 + 0w_5 + 38w_6 + 0.537w_7 + 34w_8 + b) \geq 1$

Bobot *vector* (w) adalah garis vektor yang tegak lurus antara titik pusat koordinat dengan garis *hyperplane*. Bias (b) merupakan koordinat garis *relative* terhadap titik koordinat. Persamaan 2.3 merupakan persamaan untuk menghitung nilai (b), sedangkan persamaan 2.4 merupakan persamaan untuk mencari nilai (w).

$$b = -\frac{1}{2}(w \cdot x^+ + w \cdot x^-) \quad (2.3)$$

$$w = \sum_{i=1}^n a_i y_i x_i \quad (2.4)$$

Keterangan:

b = Nilai bias.

$w \cdot x^+$ = Nilai bobot untuk kelas data positif.

$w \cdot x^-$ = Nilai bobot untuk kelas data negative.

w = Bobot vektor.

a_i = Nilai bobot data ke-i.

x_i = Data ke-i.

y_i = Kelas data ke-i.

Sebagai contoh dari persamaan sebelumnya, dengan menggunakan operasi baris elementer (OBE) dapat diperoleh nilai $w_1, w_2, w_3, w_4, w_5, w_6, w_7, w_8$ dan b sebagai berikut:

$$\begin{pmatrix} 6 & 148 & 72 & 35 & 0 & 33.6 & 0.627 & 50 & 1 \\ -1 & -85 & -66 & -29 & 0 & -26.6 & -0.351 & -31 & -1 \\ -5 & -116 & -74 & 0 & 0 & -25.6 & -0.201 & -30 & -1 \\ 3 & 78 & 50 & 32 & 88 & 31 & 0.248 & 26 & 1 \\ -10 & -115 & 0 & 0 & 0 & -35.3 & -0.134 & -29 & -1 \\ 2 & 197 & 70 & 45 & 543 & 30.5 & 0.158 & 53 & 1 \\ 8 & 125 & 96 & 0 & 0 & 0 & 0.232 & 54 & 1 \\ -4 & -110 & -92 & 0 & 0 & -37.6 & -0.191 & -30 & -1 \\ 10 & 168 & 74 & 0 & 0 & 38 & 0.537 & 34 & 1 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \\ w_6 \\ w_7 \\ w_8 \\ b \end{pmatrix}$$

$$= \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \\ w_6 \\ w_7 \\ w_8 \\ b \end{pmatrix} = \begin{pmatrix} 0.348269 \\ -0.010414 \\ 0.016492 \\ 0.023365 \\ 0.023585 \\ -0.029709 \\ 3.748240 \\ -0.007944 \\ -2.508195 \end{pmatrix}$$

Diperoleh $w_1 = 0.348269$, $w_2 = -0.010414$, $w_3 = 0.016492$, $w_4 = 0.023365$, $w_5 = 0.023585$, $w_6 = -0.029709$, $w_7 = 3.748240$, $w_8 = -0.007944$ dan $b = -2.508195$.

Data pada suatu dataset diberikan variabel x_i , sedangkan untuk kelas pada dataset diberikan variabel y_i . Metode SVM membagi dataset menjadi 2 kelas. Kelas pertama yang dipisah oleh *hyperplane* bernilai 1, sedangkan kelas lainnya bernilai -1.

$$x_i \cdot W + b \geq 1 \text{ untuk } y_i = 1 \quad (2.1)$$

$$x_i \cdot W + b \leq -1 \text{ untuk } y_i = -1 \quad (2.2)$$

Keterangan :

x_i = Data ke-i.

W = Nilai bobot *support vector* yang tegak lurus dengan *hyperlane*.

b = Nilai bias.

y_i = Kelas data ke-i.

Pada contoh sebelumnya telah dicari nilai bobot dan bias sehingga diperoleh $w_1 = 0.348269$, $w_2 = -0.010414$, $w_3 = 0.016492$, $w_4 = 0.023365$, $w_5 = 0.023585$, $w_6 = -0.029709$, $w_7 = 3.748240$, $w_8 = -0.007944$ dan $b = -2.508195$ Sehingga persamaan hyperplanenya adalah:

$$0.348269x_1 = 0.010414x_2 - 0.016492x_3 - 0.023365x_4 - 0.023585x_5 + 0.029709x_6 - 3.748240x_7 + 0.007944x_8 + 2.508195$$

Kemudian dilakukan klasifikasi data uji dengan fungsi $f(x) = -0.348269x_1 + 0.010414x_2 - 0.016492x_3 - 0.023365x_4 - 0.023585x_5 + 0.029709x_6 - 3.748240x_7 + 0.007944x_8 + 2.508195$ dengan kelasnya adalah $\text{sign}(f(x))$.

Tabel 2. Hasil klasifikasi data uji

No	Data Uji								Hasil y	Hasil Klasifikasi
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8		$sign(f(x))$
1	2	197	70	45	543	30.5	0.158	53	2.0641	+1
2	5	116	74	0	0	25.6	0.201	30	-2.75935	-1
3	10	168	74	0	0	38	0.537	34	1.815683	+1

Berikut adalah daftar karakteristik svm, seperti dikemukakan oleh Nugroho, et al., (2003).

1. SVM berfungsi sebagai classifier linier pada umumnya.
2. Untuk mengenali pola, data dari ruang input diubah menjadi ruang berdimensi lebih tinggi (ruang fitur) dan dioptimalkan dalam ruang vektor baru. Ini membedakan SVM dari teknik pengenalan pola konvensional, yang mengoptimalkan parameter pada hasil transformasi dengan dimensi lebih kecil dari ruang input.
3. Mempraktikkan rencana *Structural Risk Minimization* (SRM).
4. Pengenalan pola telah dikembangkan untuk menangani lebih dari dua jenis klasifikasi, meskipun prinsip kerja dasar SVM hanya dapat mengelola dua jenis klasifikasi.

2.4 Percentage Split

Untuk melakukan klasifikasi, data latih dan uji dataset dibagi dengan persentase. Prosedur pelatihan menggunakan persentase ini, dan data yang tersisa akan diuji. Berdasarkan seberapa berhasil mereka memperkirakan proporsi data uji yang benar, pengklasifikasi dinilai. Nilai yang dimasukkan atau dipilih menentukan berapa banyak data yang digunakan. Salah satu kemungkinan dalam opsi tes adalah pembagian persentase. Tergantung pada nilai persentase yang dimasukkan ke dalam kolom, jumlah data akan berubah (Kirkby dan Frank, 2004).

2.5 Confusion Matrix

Pengukuran terhadap performa kinerja suatu sistem klasifikasi adalah hal yang penting. Performa sistem klasifikasi menggambarkan seberapa baik sistem dapat mengklasifikasikan data. Menurut Prasetyo (2014), umumnya cara mengukur kinerja klasifikasi menggunakan *confusion matrix*. *Confusion matrix* adalah suatu metode yang dapat digunakan untuk mengukur performa dari suatu metode klasifikasi. *Confusion matrix* membandingkan hasil klasifikasi yang dilakukan oleh sistem untuk diukur keakuratannya. Selain itu menurut Rahman et., al (2017) *confusion matrix* juga dapat melakukan visualisasi terhadap hasil pembelajaran sistem, visualisasi yang ditampilkan memuat dua kategori atau lebih. Berikut adalah gambar representasi hasil proses klasifikasi *confusion matrix*.

		Actual Values	
		1 (Positive)	0 (Negative)
Predicted Values	1 (Positive)	TP (True Positive)	FP (False Positive) <i>Type I Error</i>
	0 (Negative)	FN (False Negative) <i>Type II Error</i>	TN (True Negative)

Gambar 2. Representasi hasil *confusion matrix*

Keterangan :

True Positive (TP) : Data positif yang diprediksi benar.

False Positive (FP) : Data negatif namun diprediksi sebagai data positif.

True Negative (TN) : Data negatif yang diprediksi benar.

False Negative (FN) : Data positif namun diprediksi sebagai data negatif.

Untuk menghitung nilai *precision*, *recall*, *accuracy* dan *F1-Score* dapat dilakukan dengan memasukkan data *training* ke dalam *confusion matrix*.

1. *Precision* adalah tingkat kesesuaian antara output dari sistem dengan informasi yang diperlukan.
2. *Recall* adalah tingkat keberhasilan sistem dalam mencari informasi.

3. *Accuracy* adalah tingkat keseimbangan antara nilai aktual dari nilai estimasi.
4. *F1-Score* adalah perbandingan rata-rata *precision* dan *recall* yang dibobotkan.

Precision, *recall*, *accuracy*, dan *F1-Score* secara matematis dapat dituliskan sebagai berikut:

$$Precision = \frac{TP}{TP+FP} \quad (2.8)$$

$$Recall = \frac{TP}{TP+FN} \quad (2.9)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \quad (2.10)$$

$$F1-Score = 2 \times \frac{precision \times recall}{precision+recall} \quad (2.11)$$

Keterangan:

True Positive (TP) : Data positif yang diprediksi benar.

False Positive (FP) : Data negatif namun diprediksi sebagai data positif.

True Negative (TN) : Data negatif yang diprediksi benar.

False Negative (FN) : Data positif namun diprediksi sebagai data negatif.

III.METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilaksanakan pada semester ganjil tahun ajaran 2022/2023 bertempat di Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung.

3.2 Data Penelitian

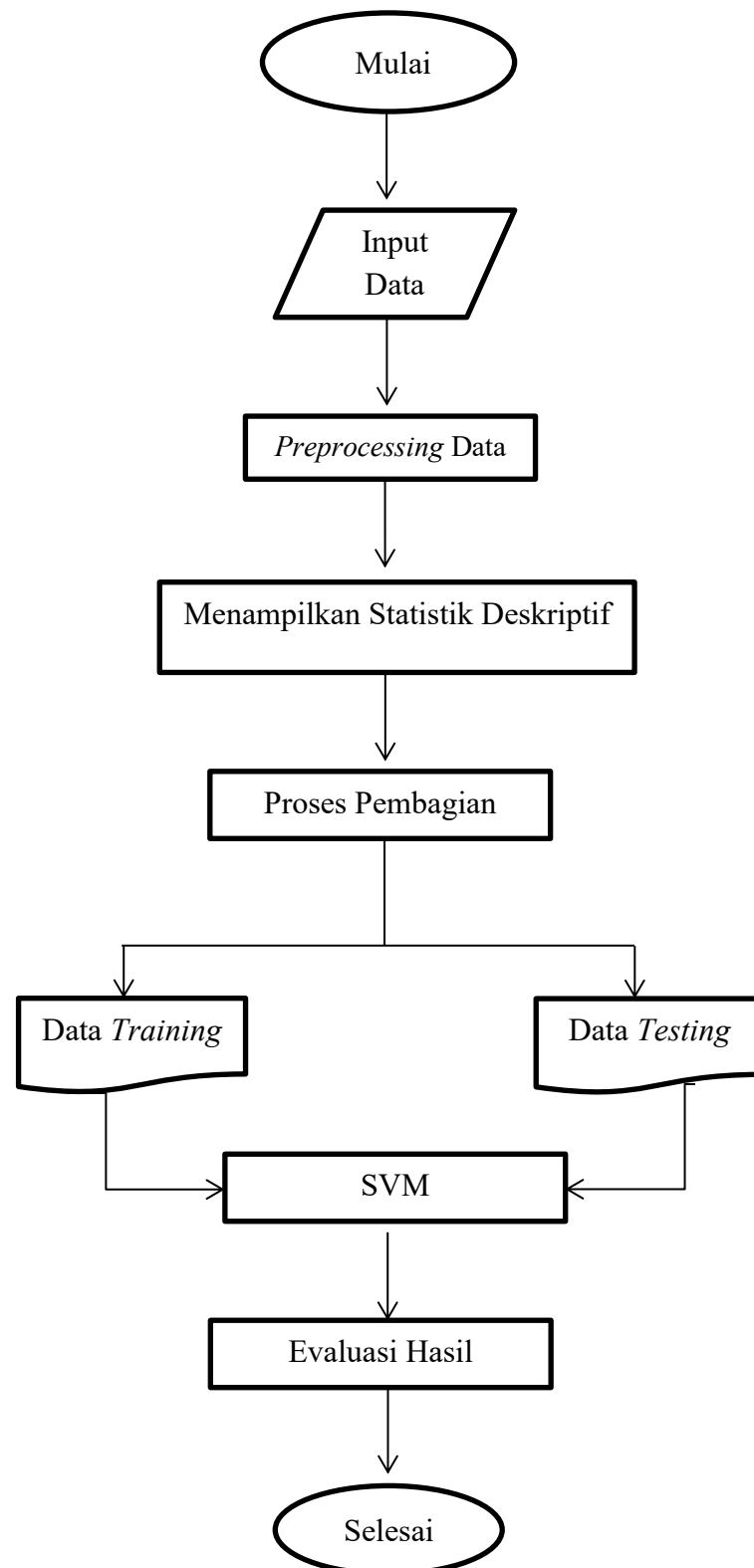
Data yang digunakan dalam penelitian ini merupakan data sekunder berupa data penderita diabetes yang diperoleh dari *Kaggle Repository* pada artikel dataset *The National Institute of Diabetes and Digestive and Kidney Disease*. Secara khusus, semua pasien disini adalah perempuan setidaknya berusia 21 tahun. Pada data ini terdapat delapan variabel independen yaitu kehamilan, glukosa, tekanan darah, ketebalan kulit, insulin, BMI, diabetes silsilah fungsi, Usia, dan satu variabel dependen yaitu status penderita diabetes.

3.3 Metode Penelitian

Pada penelitian ini akan dicari klasifikasi penyakit diabetes pada perempuan dengan bantuan *software* Phyton. Berikut adalah tahapan-tahapan yang dilakukan dalam penelitian:

1. Memasukkan syntax untuk import data ke python.
2. Melakukan pembersihan data atau *cleansing* data dengan cara menghapus *outlier* dan *missing value*.
3. Tahapan Penelitian *Support Vector Machine*:
 - a.) Menampilkan statistik deskriptif dari data.
 - b.) Eksplor data untuk mengetahui variabel independen dan nama target.
 - c.) Memeriksa dimensi data.
 - d.) Membagi dataset menjadi data *training* dan data *testing*. Adapun pembagiannya dilakukan sebanyak tiga kali sebagai berikut data *training* 80% dan data *testing* 20%, data *training* 90% dan data 10%, data *training* 95% dan data *testing* 5%.
 - e.) Input syntax untuk mendapatkan model yang akan kita gunakan untuk melakukan klasifikasi.
 - f.) Melakukan *training* untuk data *train*.
 - g.) Menampilkan *confusion matrix* untuk memudahkan dalam mengetahui apakah terdapat kesalahan dalam pengklasifikasian.
 - h.) Melihat akurasi dari hasil pengklasifikasian menggunakan algoritma SVM.

Dibawah ini merupakan *flowchart* dari metode *support vector machine*.



Gambar 3. *Flowchart* Algoritma *Support Vector Machine* (SVM).

V. PENUTUP

5.1 Kesimpulan

Berdasarkan hasil dan pembahasan yang telah diuraikan pada bab sebelumnya, maka dapat disimpulkan analisis SVM pada dataset *The National Institute of Diabetes and Digestive and Kidney Disease* diperoleh nilai *f1-score* untuk klasifikasi data negatif sebesar 0.91 dan *f1-score* untuk klasifikasi data positif sebesar 0.76. Serta, akurasi terbaik sebesar 87% dengan menggunakan kernel linear pada pembagian data *train* 95% dan data *test* 5% yang artinya penelitian ini dapat digunakan oleh masyarakat khususnya perempuan untuk antisipasi penyakit diabetes lebih dini.

5.2 Saran

Berikut adalah saran yang dibuat untuk pengembangan penelitian selanjutnya:

1. Pembuatan penelitian tambahan untuk memungkinkan perbandingan dengan algoritma lain dan penggunaan instrumen pengujian dengan nilai yang lebih tinggi.
2. Bagi peneliti yang ingin melanjutkan penelitian ini, variabel baru dapat ditambahkan atau variabel yang sudah ada dapat diubah.
3. Menyertakan gagasan tentang koefisien korelasi untuk menilai bagaimana suatu fitur memengaruhi data.

DAFTAR PUSTAKA

- Clinic, M. 2018. High Blood Pressure. <https://www.mayoclinic.org/diseases-conditions/high-bloodpressure/symptoms-causes/syc-20373410>.
Diakses pada 13 Oktober 2022.
- Dendry, Z.M., Santoso, E., & Furqon, H. Penerapan Metode Neighbor Weighted K-Nearest Neighbor Dalam Klasifikasi Diabetes Mellitus. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*. **3**(3): 2910-2915.
- Devi, Efendi, & Indriati, B. Efektivitas Air Rebusan Daun Sirih Merah Terhadap Penurunan Kadar Gula Darah Pada Pasien Diabetes Melitus Di Wilayah Kerja Puskesmas Saling. *Jurnal Keperawatan Muhammadiyah Bengkulu*. **7**(2): 559-567.
- Favorisen, R.L., Fanni, L., Yunda, H., & Kurnia, M. Implementasi SVM untuk Klasifikasi Penderita Diabetes Melitus. *Jurnal FMIPA UNILA*. **10**(1): 20-22.
- Fernanda, S.I., Ratnawati, D.E., & Adikara, P. 2017. Identifikasi Penyakit Diabetes Mellitus Menggunakan Metode Modified K-Nearest Neighbor (MKNN). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*. **1**(-): 507-513.
- Han, J. dan Kamber M. 2006. Data Mining Concepts and Techniques. 2th Edition. Morgan Kaufmann, San Francisco.

- Kirkby & Frank. 2004. WEKA Explorer User Guide for Version 3-4-3. University of Waikato.
- Nugroho. 2008. Support Vector Machine Paradigma Baru Dalam Softcomputing, hlm 92-99. Konferensi Nasional sistem dan Informatika, Jakarta.
- Nugroho, A.S., Witarto, B.A., Handoko, D. 2003. Application of Support Vector Machine in Bioinformatics, hlm. 88-94. Proceeding of Indonesian Scientific Meeting in Central Japan, Gifu.
- Prasetyo, E. 2014. Data Mining : Konsep dan Aplikasi Menggunakan MATLAB. Andi Offset, Yogyakarta.
- Rachman, 2012. Perbandingan Klasifikasi Tingkat Keganasan Breast Cancer dengan Menggunakan Regresi Logistik Ordinal dan Support Vector Machine. *Jurnal Sains dan Seni ITS*. **1**(1): 29-33.
- Rahman, M.F., Darmawidjaja, M.I., & Alamsah, D. 2017. Klasifikasi untuk Diagnosa Diabetes Menggunakan Metode Bayesan Regularization Neural Network (RBNN). *Jurnal Informatika*. **11**(1): 36-45.
- Raudhatul. 2016. Penerapan Metode Support Vector Machine Pada Diagnosa Hepatitis. *Jurnal Nasional Komputasi dan Teknologi Informasi*. **5**(3): 442-449.
- Siagian, R. 2011. Klasifikasi Parket Kayu Jati Menggunakan Metode Support Vector Machine (SVM). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*. **2**(2): 802-810.