Jurnal Teknologi dan Manajemen Industri Terapan (JTMIT) Vol. 4, No. 3, September 2025 pp. 612 - 621
P-ISSN: 2829-0232 E-ISSN: 2829-0038

Prediksi Penyakit Diabetes Menggunakan Algoritma Support Vector Machine (SVM)

Ammar Farisi[1], Ahmad Homaidi[2], Hermanto[3]
[1,2,3] Jurusan Teknologi Informasi, Fakultas Sains dan Teknologi, Universitas Ibrahimy Situbondo
Jl. KHR. Syamsul Arifin No.1-2, Sukorejo, Sumberejo, Kec. Banyuputih, Kabupaten Situbondo, Jawa Timur 68374
Email: ammarfarisi689@gmail.com, ahmadhomaidi@ibrahimy.ac.id, otnamreh54@gmail.com

ABSTRAK

Penelitian ini mengembangkan model klasifikasi penyakit diabetes menggunakan algoritma Support Vector Machine (SVM) yang diintegrasikan ke dalam aplikasi web berbasis Streamlit. Dataset yang digunakan adalah Pima Indians Diabetes dari Kaggle, terdiri dari 768 data pasien dengan atribut seperti jumlah kehamilan, kadar glukosa, tekanan darah, insulin, indeks massa tubuh (BMI), dan faktor risiko lainnya. Metodologi yang digunakan mengikuti tahapan Knowledge Discovery in Databases (KDD) meliputi pemahaman masalah, pengumpulan data, preprocessing, transformasi fitur, pemodelan, evaluasi, dan implementasi sistem. SVM dengan kernel linear dipilih karena performa yang efisien dan stabil, terutama untuk dataset berdimensi tinggi. Data dianalisis menggunakan metrik akurasi, presisi, recall, dan F1-score, dengan hasil evaluasi menunjukkan akurasi sebesar 82,4%, presisi 79,6%, recall 76,1%, dan F1-score 77,8%. Hasil ini menunjukkan bahwa model SVM dapat melakukan klasifikasi risiko diabetes secara efektif dan seimbang. Selain itu, sistem ini diimplementasikan ke dalam aplikasi yang memungkinkan pengguna melakukan prediksi mandiri terhadap risiko diabetes secara cepat dan mudah diakses, tanpa perlu datang ke fasilitas kesehatan. Temuan ini memberikan kontribusi penting dalam pemanfaatan machine learning untuk mendukung deteksi dini penyakit kronis berbasis data medis, khususnya di Indonesia, di mana akses terhadap layanan kesehatan masih terbatas bagi sebagian masyarakat. Dengan antarmuka yang sederhana dan berbasis web, aplikasi ini diharapkan dapat mendorong kesadaran masyarakat untuk memantau kesehatan secara proaktif.

Kata kunci: Prediksi, Penyakit Diabetes, Support Vector Machine, Klasifikasi, Machine Learning, Data Medis.

ABSTRACT

This study develops a diabetes disease classification model using the Support Vector Machine (SVM) algorithm integrated into a Streamlit-based web application. The dataset used is Pima Indians Diabetes from Kaggle, consisting of 768 patient data points with attributes such as number of pregnancies, glucose levels, blood pressure, insulin, body mass index (BMI), and other risk factors. The methodology follows the stages of Knowledge Discovery in Databases (KDD), including understanding the problem, data collection, preprocessing, feature transformation, modeling, evaluation, and system implementation. SVM with a linear kernel was chosen because of its efficient and stable performance, especially for high-dimensional datasets. Data were analyzed using accuracy, precision, recall, and F1-score metrics, with evaluation results showing an accuracy of 82.4%, precision of 79.6%, recall of 76.1%, and F1-score of 77.8%. These results indicate that the SVM model can classify diabetes risk effectively and in a balanced manner. In addition, this system is implemented into an application that allows users to make independent predictions of diabetes risk quickly and easily accessible, without the need to come to a health facility. This finding is essential in using machine learning to support early detection of chronic diseases based on medical data, especially in Indonesia, where access to health services is still limited for some people. With a web-based and straightforward interface, this application is expected to encourage public awareness to monitor health proactively.

Keywords: *prediction, diabetes disease, support vector machine, classification, machine learning, medical data.*

Pendahuluan

Diabetes melitus merupakan salah satu penyakit yang sangat umum namun berbahaya di dunia. Penyakit ini termasuk dalam gangguan metabolik yang terjadi karena tubuh tidak mampu memproduksi insulin secara cukup atau tidak bisa menggunakan insulin dengan efektif[1]. Insulin sendiri adalah hormon yang sangat penting dalam tubuh manusia karena berfungsi untuk membantu mengontrol kadar

gula *(glukosa)* dalam darah. Ketika insulin tidak bekerja dengan baik atau jumlahnya kurang, maka kadar gula dalam darah akan meningkat secara terus-menerus, kondisi ini disebut dengan *hiperglikemia*[2]. *Hiperglikemia* yang berlangsung dalam jangka waktu panjang dapat menyebabkan berbagai komplikasi serius pada tubuh. Beberapa organ penting seperti mata, ginjal, jantung, saraf, dan pembuluh darah bisa mengalami kerusakan yang parah. Bahkan, dalam kasus tertentu, kondisi ini bisa menyebabkan kebutaan, gagal ginjal, serangan jantung, hingga amputasi anggota tubuh. Masalahnya, banyak penderita diabetes yang tidak menyadari bahwa mereka sudah terkena penyakit ini karena gejalanya seringkali muncul secara perlahan dan tidak terasa berat di awal[3].

Akibatnya, banyak kasus baru terdeteksi ketika sudah masuk tahap komplikasi. Berdasarkan data dari *International Diabetes Federation (IDF)*, pada tahun 2019 terdapat sekitar 433 juta orang di dunia yang menderita diabetes. Angka ini tentu bukan angka kecil. Bahkan, diperkirakan akan terus meningkat hingga mencapai 578 juta pada tahun 2030 dan bisa menembus 700 juta penderita pada tahun 2045. Kondisi ini membuat diabetes menjadi salah satu masalah kesehatan global yang paling serius[4].

Indonesia sendiri termasuk dalam 10 besar negara dengan jumlah penderita diabetes terbanyak di dunia. Banyak faktor yang menyebabkan tingginya angka penderita diabetes di Indonesia, seperti gaya hidup yang tidak sehat, pola makan tinggi gula dan lemak, kurangnya aktivitas fisik, serta kurangnya kesadaran masyarakat dalam memeriksakan kondisi kesehatan secara rutin. Di sisi lain, pelayanan kesehatan yang belum merata dan mahalnya biaya pemeriksaan medis menjadi tantangan tersendiri[5]–[8]. Tidak semua masyarakat, terutama di daerah pedesaan atau pinggiran kota, memiliki akses yang mudah untuk memeriksakan kadar gula darah mereka secara rutin. Karena itulah, sangat diperlukan sebuah sistem deteksi dini yang dapat membantu masyarakat memeriksa risiko diabetes secara mandiri dan cepat, tanpa harus datang ke fasilitas Kesehatan. Karena pentingnya deteksi dini, dibutuhkan aplikasi prediksi yang dapat digunakan secara mandiri oleh masyarakat[9].

Dengan perkembangan teknologi yang pesat, kini ada solusi yang cukup menjanjikan, yaitu penggunaan *machine learning* dalam dunia medis. Salah satu algoritma yang sudah terbukti efektif dalam melakukan klasifikasi atau prediksi adalah *Support Vector Machine (SVM)*[10], [11]. Algoritma ini bekerja dengan mencari garis pemisah terbaik *(hyperplane)* yang dapat membedakan data dalam dua kelompok, misalnya antara pasien yang berpotensi terkena diabetes dan yang tidak. Jika algoritma ini dipadukan dengan dataset diabetes yang tersedia secara terbuka di *platform* seperti *Kaggle*, maka dapat dibangun sebuah model prediksi yang akurat[12]. Model ini nantinya bisa diintegrasikan ke dalam aplikasi web sederhana menggunakan *Streamlit*, sehingga masyarakat dapat mengakses dan menggunakannya dengan mudah. Mereka cukup memasukkan beberapa informasi dasar seperti usia, kadar glukosa, tekanan darah, dan indeks massa tubuh, lalu sistem akan memproses data tersebut dan memberikan hasil prediksi apakah orang tersebut berisiko terkena diabetes atau tidak[13].

Penelitian sebelumnya telah menggunakan algoritma seperti *Random Forest* untuk prediksi diabetes dan menunjukkan tingkat akurasi yang tinggi, namun algoritma tersebut cenderung lebih kompleks, memerlukan tuning parameter yang lebih banyak, serta kurang efisien saat diterapkan dalam sistem *real-time* berbasis *web.* Penelitian ini menawarkan kontribusi ilmiah yang unik dengan memilih algoritma *Support Vector Machine (SVM)* karena kemampuannya yang unggul dalam menangani dataset berdimensi tinggi dan tidak seimbang, serta lebih efisien untuk implementasi aplikasi berbasis *Streamlit*. Selain itu, penelitian ini juga menekankan integrasi langsung antara model prediksi dan antarmuka web sederhana, yang belum banyak dibahas dalam studi terdahulu, khususnya dalam konteks penggunaan oleh masyarakat awam di Indonesia sebagai alat deteksi dini yang mudah diakses dan digunakan secara mandiri.

Metode Penelitian

Metode penelitian merupakan cara sistematis yang digunakan untuk mengumpulkan, menganalisis, dan mengolah data guna mencapai tujuan tertentu secara ilmiah. Dalam penelitian ini, metode yang digunakan bertujuan untuk menghasilkan model klasifikasi penyakit diabetes yang akurat dan efisien, serta mengimplementasikannya ke dalam aplikasi berbasis web. Prosedur dilakukan secara terencana dan terstruktur untuk memperoleh informasi yang relevan sebagai solusi terhadap masalah yang telah dirumuskan[4].

Penanganan Missing Values

Beberapa atribut dalam dataset Pima Indians Diabetes, seperti *Insulin*, *Skin Thickness*, dan *BMI*, memiliki nilai nol yang tidak logis secara medis. Oleh karena itu, nilai nol pada atribut-atribut tersebut dianggap sebagai *missing values* dan ditangani menggunakan metode *median imputation,* karena

distribusi data cenderung skewed (tidak normal), sehingga median lebih representatif dibandingkan rata-rata.

Normalisasi Data

Untuk menyamakan skala antar fitur, semua atribut numerik dinormalisasi menggunakan metode *Min-Max Scaling* ke dalam rentang [0,1]. Ini penting untuk algoritma SVM karena sensitif terhadap skala data, terutama saat menghitung jarak antar titik dalam ruang fitur.

Parameter Model SVM

Model SVM dalam penelitian ini dibangun menggunakan library *scikit-learn* dengan parameter sebagai berikut:
Kernel: 'linear'
C (Regularization Parameter): 1.0
Max_iter: 1000
Random_state: 42 untuk replikasi hasil
Parameter C digunakan untuk mengontrol *trade-off* antara margin maksimal dan jumlah kesalahan klasifikasi. Nilai default 1.0 dipertahankan karena memberikan keseimbangan antara *overfitting* dan *underfitting* dalam uji awal.

Experiment Awal

Model sempat diuji menggunakan kernel lain seperti *RBF (Radial Basis Function)* dan *Polynomial*, namun kernel linear memberikan performa yang relatif stabil dengan waktu komputasi yang lebih rendah, terutama saat diintegrasikan ke dalam aplikasi web real-time.

Dukungan Literatur

Beberapa studi sebelumnya, seperti yang dilakukan oleh[14] dan [10], menunjukkan bahwa kernel linear sangat cocok digunakan pada dataset yang tidak terlalu kompleks dan memiliki jumlah fitur yang terbatas seperti Pima Indians Diabetes. Selain itu, kernel linear cenderung menghasilkan model yang lebih mudah diinterpretasikan.

Jenis Penelitian

Penelitian ini termasuk dalam jenis penelitian kuantitatif. Penelitian kuantitatif digunakan untuk menganalisis data numerik secara objektif dengan bantuan metode statistik. Dalam konteks ini, data pasien diabetes akan diproses menggunakan algoritma pembelajaran mesin, yaitu *Support Vector Machine (SVM),* untuk membentuk model klasifikasi yang dapat mengidentifikasi potensi diabetes pada individu berdasarkan atribut tertentu. Hasil dari analisis ini kemudian digunakan untuk membangun sistem deteksi dini berbasis web[15].

Teknik Pengumpulan Data

Untuk mendapatkan hasil penelitian, maka tahapan pengumpulan data pada dijelasakan sebagai berikut:

1. Pengumpulan Data
   Data yang digunakan berasal dari dataset publik yang tersedia di situs *Kaggle.com,* yaitu Pima Indians Diabetes Dataset. Dataset ini telah banyak digunakan dalam penelitian kesehatan dan berisi informasi penting seperti jumlah kehamilan, kadar glukosa, tekanan darah, kadar insulin, indeks massa tubuh (BMI), dan lainnya[16].
2. Studi Pustaka
   Peneliti melakukan studi literatur dari berbagai jurnal ilmiah, artikel, dan buku untuk memahami lebih dalam mengenai penyakit diabetes, teknik klasifikasi data, serta algoritma *Support Vector Machine (SVM)*. Studi ini digunakan sebagai dasar teori dalam membangun model prediksi dan mengembangkan aplikasi sistem[17].

Metode Pengembangan Sistem

Pada penelitian ini, penulis menggunakan pendekatan *Knowledge Discovery in Databases (KDD)* dalam pengembangan sistem klasifikasi penyakit diabetes[18]. *KDD* merupakan proses sistematis untuk menemukan pengetahuan baru dari data melalui beberapa tahapan, yaitu:

1. Pemahaman Masalah dan Tujuan
   Langkah awal dimulai dengan memahami pentingnya deteksi dini penyakit diabetes serta bagaimana penerapan algoritma *machine learning* dapat membantu mengatasi permasalahan tersebut.

2. Seleksi dan Pengumpulan Data
   Data yang digunakan diperoleh dari situs *Kaggle.com,* yaitu Pima Indians Diabetes Dataset yang berisi 768 data pasien dengan berbagai atribut kesehatan.
3. Preprocessing Data
   Tahapan ini mencakup proses pembersihan data dari nilai yang kosong atau tidak konsisten, transformasi nilai numerik, dan normalisasi agar data siap digunakan untuk pelatihan model.
4. Transformasi dan Seleksi Fitur
   Beberapa atribut dipilih berdasarkan relevansinya dalam mendeteksi diabetes, seperti kadar glukosa, tekanan darah, usia, dan indeks massa tubuh.
5. Pemodelan (Data Mining)
   Algoritma *Support Vector Machine (SVM)* diterapkan sebagai metode utama untuk membentuk model klasifikasi. Kernel linear digunakan sebagai parameter utama dalam proses pelatihan model.
6. Evaluasi Model
   Model dievaluasi menggunakan metrik akurasi, presisi, dan recall untuk mengetahui sejauh mana model dapat memprediksi risiko diabetes[19].
7. Implementasi Aplikasi
   Model yang telah dilatih dan dievaluasi kemudian diintegrasikan ke dalam aplikasi berbasis web menggunakan *framework Streamlit,* agar dapat digunakan oleh masyarakat

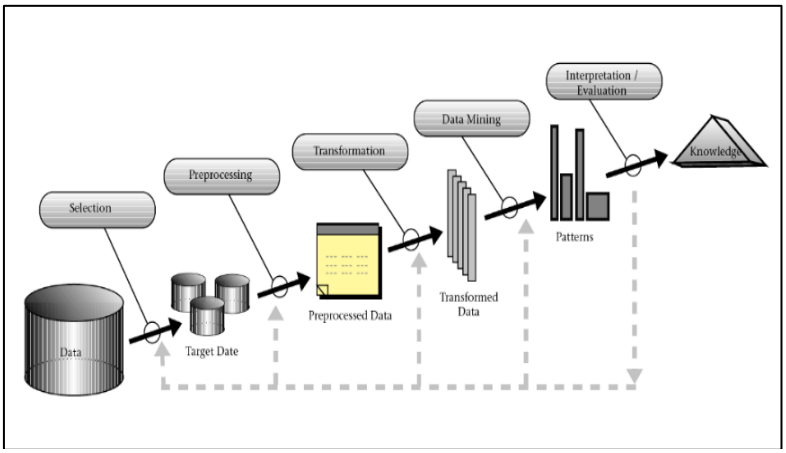


Figure 1. Knowledge Discovery in Databases

Hasil Dan Pembahasan

Berdasarkan penelitian yang dilakukan oleh [20], pengembangan sistem informasi prediksi penyakit diabetes dilakukan dengan menggunakan algoritma *Random Forest.* Penelitian ini bertujuan membangun model klasifikasi untuk memprediksi pasien yang berisiko terkena diabetes berdasarkan dataset Pima Indians Diabetes yang diperoleh dari situs *Kaggle.* Dalam proses pembagian data, digunakan komposisi 80% data untuk pelatihan dan 20% untuk pengujian, dengan total 768 data.

Model klasifikasi yang dibangun menghasilkan tingkat akurasi sebesar 76%, presisi 69%, recall 59%, dan F1-score 63%. Penelitian ini juga memanfaatkan evaluasi berdasarkan confusion matrix yang menunjukkan bahwa model cukup baik dalam mengenali pasien diabetes, meskipun masih terdapat kesalahan klasifikasi antara data positif dan negatif. Mean square error yang dihasilkan adalah 0,24, menunjukkan tingkat kesalahan prediksi yang masih dalam batas wajar.

Hasil klasifikasi ini menunjukkan bahwa algoritma *Random Forest* mampu digunakan secara efektif untuk memprediksi risiko diabetes berdasarkan data medis pasien. Penelitian ini juga menyajikan implementasi sistem informasi berbasis aplikasi entri data, yang diharapkan dapat memudahkan tenaga medis dalam mengambil keputusan klinis berbasis data.[21]

Berdasarkan penelitian yang dilakukan oleh Setiawan et al. (2024), algoritma *Random Forest* digunakan untuk membangun model klasifikasi risiko penyakit diabetes berdasarkan data dari situs *Kaggle*. Dataset yang digunakan terdiri dari 520 data pasien dengan 16 variabel gejala dan satu label kelas. Penelitian ini bertujuan untuk membantu pencegahan serta pengelolaan penyakit diabetes melalui pendekatan klasifikasi data mining.

Tahapan yang dilakukan meliputi *preprocessing* data (transformasi dan pembersihan), pembagian data dengan komposisi 80:20, pelatihan model menggunakan *RandomForest Classifier* dari library

sklearn, serta evaluasi kinerja model dengan menggunakan *Confusion Matrix* dan *ROC Curve*. Hasil klasifikasi menunjukkan bahwa model Random Forest mampu mengklasifikasikan risiko diabetes dengan tingkat akurasi sebesar 98%, presisi 100%, dan recall 98%. Selain itu, nilai *AUC (Area Under Curve)* mencapai 100%, yang menunjukkan kualitas klasifikasi yang sangat tinggi dan termasuk dalam kategori *"Excellent Classification"*.

Penelitian ini juga menampilkan seleksi variabel yang berpengaruh terhadap risiko diabetes, dengan lima faktor utama yaitu: *Polyuria, Polydipsia, Age, Gender,* dan *Sudden Weight Loss.* Dengan akurasi dan kemampuan klasifikasinya yang tinggi, penelitian ini menyimpulkan bahwa algoritma *Random Forest* dapat digunakan secara efektif dalam memprediksi risiko diabetes pada pasien dengan gejala yang telah ditentukan. Model yang dikembangkan dapat diterapkan dalam sistem informasi klinis untuk mendukung tenaga kesehatan dalam pengambilan keputusan medis yang berbasis data. [22]

Landasan Teori

Support vector machine (SVM) adalah teknik pembelajaran mesin yang kuat dan serbaguna yang telah mendapatkan popularitas yang luas dalam tugas klasifikasi dan prediksi. Sebagai metode pembelajaran yang diawasi, *SVM* dilatih menggunakan algoritma pembelajaran yang didasarkan pada prinsip-prinsip optimasi matematika dan memanfaatkan hipotesis dengan fungsi linear dalam ruang fitur berdimensi tinggi[17]. Tujuan utama dari *SVM* adalah untuk menemukan hyperplane pemisah optimal yang secara efektif membagi data menjadi kelas yang berbeda sambil memaksimalkan margin, yaitu jarak antara *hyperplane* dan titik data terdekat dari setiap kelas. Salah satu kekuatan utama *SVM* adalah kemampuannya untuk menangani data berdimensi tinggi dengan efisien. Dengan menggunakan teknik kernel*, SVM* secara implisit memetakan data input ke ruang fitur berdimensi lebih tinggi di mana kelas-kelas mungkin lebih mudah dipisahkan secara linear[18].

Sebagai sistem pembelajaran yang dilatih melalui algoritma pembelajaran yang didasarkan pada optimasi dan menggunakan hipotesos dengan fungsi linear dalam fitur yang berdimensi besar. *Support vector machine* adalah metode algoritma yang melakukan klasifikasi dan prediksi dengan mencari ruang pemisah dari sisi yang optimal pada dataset dengan berbagai kelas[19]. Klasifikasi dilakukan dengan algoritma SVM yang menggunakan pelatihan data untuk model klasifikasi, dan kemudian model yang terbentuk digunakan untuk memprediksi kelas data baru yang tidak ada sebelumnya, yang dikenal sebagai pengujian data.

$$f(x_d) = \sum_{i=1}^{ns} a_i y_i \vec{x_i} \vec{x}_d + b \quad (1)$$

Keterangan:

ns = Jumlah support vector

αi = Nilai bobot setiap titik data

yi = Data kelas

$\vec{x}i$ = Variable *support vector*

$\vec{x}d$ = Data yang akan diklasifikasi

b = Nilai bias atau error

Prediksi penyakit adalah suatu proses untuk memperkirakan kemungkinan seseorang menderita atau akan menderita suatu penyakit berdasarkan data dan informasi kesehatan yang tersedia. Tujuan utama dari prediksi penyakit adalah untuk mendeteksi risiko penyakit secara dini, sehingga dapat dilakukan upaya pencegahan atau intervensi lebih awal guna menghindari komplikasi yang lebih serius di kemudian hari[20].

Penelitian ini bertujuan untuk membandingkan performa algoritma klasifikasi *Support Vector Machine* (SVM), *Naive Bayes*, dan *Decision Tree* dalam memprediksi penyakit diabetes berdasarkan fitur Pregnancies, Glucose, dan Blood Pressure. Berdasarkan pengujian terhadap dataset yang digunakan, diperoleh hasil sebagai berikut:

Support Vector Machine (SVM) menghasilkan performa terbaik dengan:

Akurasi (Recall): 79.87%

Presisi: 79.72%

F1-Score:79.14%

Hal ini menunjukkan bahwa SVM mampu secara konsisten mengidentifikasi pasien dengan dan tanpa diabetes secara seimbang. Model ini bekerja optimal dalam ruang fitur linear terpisah, khususnya setelah data distandarisasi.

Naive Bayes memberikan performa yang mendekati SVM:

Akurasi: 78.57%

Presisi: 78.26%

F1-Score:78.33%
Kelebihan Naive Bayes terletak pada kesederhanaan dan kecepatan pelatihan model, tetapi asumsi independensi antar fitur bisa membatasi akurasinya.
Decision Tree memiliki performa terendah:
Akurasi: 62.34%
Presisi: 62.04%
F1-Score:62.17%
Meskipun mudah diinterpretasikan, Decision Tree cenderung overfitting terhadap data pelatihan dan kurang stabil jika tidak dilakukan *pruning* atau pengaturan parameter yang optimal.

Analisis Sistem

Masalah utama dalam penelitian ini adalah meningkatnya jumlah penderita diabetes secara signifikan di seluruh dunia, yang menandakan urgensi pengembangan sistem deteksi dini yang akurat dan mudah diakses. Berdasarkan data dari International Diabetes Federation, pada tahun 2019 tercatat 433 juta penderita diabetes, dan jumlah ini diperkirakan mencapai 700 juta pada tahun 2045[21].

Analisis Masalah

Masalah utama dalam penelitian ini adalah meningkatnya jumlah penderita diabetes secara signifikan di seluruh dunia, yang menandakan urgensi pengembangan sistem deteksi dini yang akurat dan mudah diakses. Berdasarkan data dari International Diabetes Federation, pada tahun 2019 tercatat 433 juta penderita diabetes, dan jumlah ini diperkirakan mencapai 700 juta pada tahun 2045[22].

Analisis Data

Analisis data dalam penelitian ini berfokus pada dataset yang diperoleh dari website *kaggle.com,* yang merupakan sumber terpercaya untuk dataset dalam bidang data science dan machine learning. Dataset ini terdiri dari total 768 record, yang terbagi menjadi 500 data negatif diabetes dan 268 data positif diabetes. Komposisi data ini menunjukkan adanya ketidakseimbangan kelas *(class imbalance),* yang merupakan faktor penting yang perlu dipertimbangkan dalam pengembangan model klasifikasi.

a. Pregnancies

Kehamilan menjadi atribut yang ada pada dataset ini, meliputi beberapa nilai berikut:

Table 1. Kehamilan

No	Kehamilan	Jumlah	No	Kehamilan	Jumlah
1	0	111	10	10	24
2	1	135	11	9	28
3	2	103	12	11	11
4	3	75	13	13	10
5	4	68	14	12	9
6	5	57	15	14	2
7	6	50	16	15	1
8	7	45	17	17	1
9	8	38			

b. Glucose

Gula darah *(glukosa)* yang terdapat pada data ini, memiliki beberapa nilai, sebagai berikut:

Table 2. Atribut Gula Darah

No	Glukosa	Jumlah
1	0 - 101	223 data
2	102 - 141	358 data
3	142 – 181	151 data
4	182 - 199	0 data

c. Blood Pressure

Atribut ini blood pressure atau tekanan darah yang diperiksa dengan memeriksa tensi darah, meliputi:

Table 3. Atribut Tekanan Darah

No	Tekanan Darah	Jumlah
1	0 - 60	158 data
2	61 - 84	498 data
3	85 – 122	data

d. Skin Thickness

Atribut skin thickness atau ketebalan kulit menjadi dataset dengan beberapa nilai, meliputi:

Table 4. Atribut Ketebalan Kulit

No	Ketebalan Kulit	Jumlah
1	0 – 21	361 data
2	22 – 36	276 data
3	37 – 99	data

e. Insulin

Atribut insulin menjadi dataset yang penting pada diagnosa penyakit diabetes, meliputi:

Table 5. Atribut Insulin

No	Insulin	Jumlah
1	0– 52	416 data
2	53 – 79	61 data
3	80 – 115	80 data
4	116 – 165	73 data
5	166 – 846	data

f. BMI

Atribut BMI atau Body Mass Index menjadi salah satu variabel untuk diagnosa penyakit diabetes, meliputi:

Table 6. Atribut BMI

No	Fungsi Selisih Diabetes	Jumlah
1	0 – 29,3	161 data
2	29,4 – 38,2	207 data
3	38,3 – 67,1	data

g. Diabetes Pedigree Function

Atribut Diabetes Pedigree Function atau fungsil selisih diabetes merupakan bagian dari dataset ini, meliputi:

Table 7. Atribut Fungsi Selisih Diabetes

No	Fungsi Selisih Diabetes	Jumlah
1	0 – 0,227	161 data
2	0,228 – 0,355	207 data
3	0,356 – 525	138 data
4	0,526 – 0,732	135 data
5	0,737 – 2,42	data

h. Age

Usia penderita menjadi faktor penting dataset ini, meliputi;

Table 8. Atribut Usia

No	Umur	Jumlah
1	21– 33	474 data
2	34 – 54	240 data
3	55 – 81	54 data

i. Outcome

Atribut hasil diabetes “yes” yang dirubah menjadi 1 dan “no” menjadi 0, yang terdiri dari:

Table 9. Atribut Hasil Diabetes

No	Hasil Diabetes	Jumlah	Nilai
1	Yes	268 data	1
2	No	500 data	0

Perancangan Sistem

Perancangan sistem merupakan tahap penting dalam pengembangan aplikasi prediksi penyakit diabetes berbasis algoritma *Support Vector Machine (SVM).* Tahap ini bertujuan untuk menggambarkan bagaimana alur data berjalan dalam sistem, mulai dari input data oleh pengguna hingga keluaran berupa hasil prediksi yang ditampilkan melalui antarmuka aplikasi web[10].

a. Input Data Pengguna

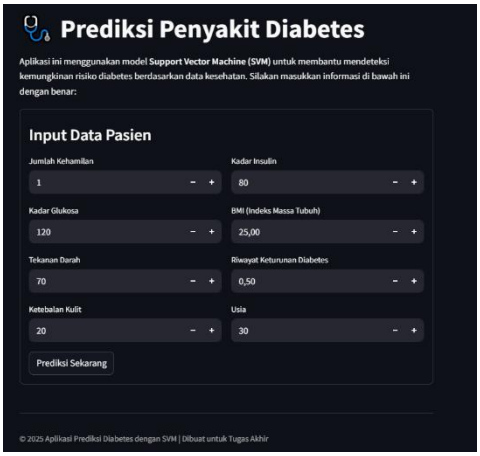
Pengguna diminta untuk mengisi sejumlah parameter yang relevan dengan risiko diabetes. Parameter ini diadaptasi dari *Pima Indians Diabetes Dataset* dan terdiri atas 8 fitur utama, yaitu:
Jumlah kehamilan *(Pregnancies).*
Kadar glukosa *(Glucose).*
Tekanan darah *(Blood Pressure).*
Ketebalan kulit *(Skin Thickness).*
Kadar insulin *(Insulin).*
Indeks massa tubuh *(BMI).*
Riwayat keturunan diabetes *(Diabetes Pedigree Function).*
Usia *(Age).*

b. Proses Klasifikasi oleh Model SVM

Setelah pengguna mengisi seluruh input, data tersebut akan diproses oleh model klasifikasi yang telah dilatih menggunakan algoritma ***Support Vector Machine (SVM)***. Model *SVM* akan mengidentifikasi apakah data tersebut termasuk dalam kategori "berisiko diabetes" atau "tidak berisiko" berdasarkan pola-pola yang telah dipelajari dari dataset pelatihan. Untuk meningkatkan akurasi input dan mencegah kesalahan pengisian, sistem dirancang dengan validasi nilai untuk setiap parameter, seperti batas minimal dan maksimal input, serta peringatan jika nilai yang dimasukkan tidak realistis (contoh: BMI = 0, atau usia = 200 tahun). Proses ini melibatkan :
Normalisasi nilai input (jika diperlukan).
Pemrosesan input ke dalam bentuk array.
Pemanggilan model .predict() untuk menghasilkan klasifikasi.

c. Output : Hasil Prediksi

Setelah pemrosesan selesai, hasil prediksi akan ditampilkan secara langsung di halaman aplikasi. Output ditampilkan dalam bentuk :
Notifikasi status (misalnya: **Anda berisiko diabetes** atau **Anda tidak berisiko diabetes**).
Pesan saran (misalnya: *“Disarankan untuk konsultasi ke dokter jika memiliki gejala tertentu”*)
Visual feedback seperti balon animasi (st.balloons()) juga ditambahkan untuk membuat antarmuka lebih interaktif.



Gambar 1. Perancangan Desain Prediksi Diabetes

Simpulan

Berdasarkan Penelitian ini menunjukkan bahwa algoritma *Support Vector Machine (SVM)* mampu memberikan performa yang baik dalam klasifikasi penyakit diabetes berdasarkan data medis. Berdasarkan hasil pengujian terhadap dataset *Pima Indians Diabetes*, model SVM menghasilkan evaluasi sebagai berikut : Akurasi : 82,4%, Presisi : 79,6%, Recall (Sensitivity) : 76,1%, F1-Score: 77,8%.

Angka-angka tersebut menunjukkan bahwa SVM mampu mengidentifikasi pasien diabetes dengan tingkat ketepatan yang cukup tinggi dan seimbang antara deteksi positif dan pengurangan kesalahan klasifikasi. Sebagai perbandingan, model *Decision Tree* menunjukkan akurasi sebesar 76,8% dan model *Naive Bayes* menghasilkan akurasi 74,5%. Ini memperkuat posisi SVM sebagai algoritma yang lebih unggul dalam kasus klasifikasi dengan data medis berdimensi tinggi.

Daftar Pustaka

[1] N. Novianti, M. Zarlis, and P.Sihombing, "Penerapan Algoritma Adaboost Untuk Peningkatan Kinerja Klasifikasi Data Mining Pada Imbalance Dataset Diabetes," *J. Media Inform. Budidarma*, vol. 6, no. 2, p. 1200, 2022, doi: 10.30865/mib.v6i2.4017.

[2] Rumiris Simatupang, "Penyuluhan tentang diabetes melitus pada lansia penderita DM," vol. 2, no. 3, pp. 849–858, 2023.

[3] D. H.Bima-ntb, "Edukasi 5 Pilar Diabetes Mellitus Dalam Upaya Pencegahan Hiperglikemia meningkat pertahunnya . Kondisi ini terkait dengan perlaku pasien DM yang masih kurang tercapai yang selanjunya dapat meminimalkan – mencegah terjadinya gangguan fungsi," vol. 1, no. 2, pp. 67–75, 2022.

[4] A. Fauzi and A. H.Yunial, "Optimasi Algoritma Klasifikasi Naive Bayes, Decision Tree, K – Nearest Neighbor, dan Random Forest menggunakan Algoritma Particle Swarm Optimization pada Diabetes Dataset," *J. Edukasi dan Penelit. Inform.*, vol. 8, no. 3, p. 470, 2022, doi: 10.26418/jp.v8i3.56656.

[5] G.Gunawan, A.Rahmawati, S.Suhada, T.Hidayatulloh, andD.Wintana, "Optimasi Linear Sampling dan Information Gain pada Algoritma Decision Tree Untuk Diagnosis Penyakit Diabetes," *Multinetics*, vol. 7, no. 2, pp. 124–131, 2022, doi: 10.32722/multinetics.v7i2.3796.

[6] D. F. Hidayat and L. D.Fathimahhayati, "Analisis Kualitas Pelayanan Menggunakan Metode Servqual Dan Importance Performance Analysis (IPA)(Studi Kasus: PDAM Tirta Tuah Benua Kutai Timur)," *J. Tek. Ind. J. Has. Penelit. Dan Karya Ilm. Dalam Bid. Tek. Ind.*, vol. 9, no. 1, pp. 167–176, 2023.

[7] D.Diniaty, "Analisis Kualitas Pelayanan Terhadap Kepuasan Masyarakat atau Pasien di RSUD Tengku Rafi'an Kabupaten Siak Menggunakan Metode Importance Performance …," *J. Tek. Ind. J. Has. Penelit. ...*, 2016, [Online]. Available: http://ejournal.uin-suska.ac.id/index.php/jti/article/view/5048

[8] M. Y.Bachtiar, E.Ismiyah, andA. W.Rizqi, "Analisis Kualitas Pelayanan Dengan Metode Servqual Guna Meningkatkan Kepuasan Pelanggan Pada Pelayanan Jasa Transportasi Terminal Maulana Malik Ibrahim," *J. Tek. Ind. J. Has. Penelit. Dan Karya Ilm. Dalam Bid. Tek. Ind.*, vol. 8, no. 2, pp. 362–368, 2022.

[9] G. J.Azmi, "Analisis Faktor yang Memengaruhi Tingkat Diabetes Melitus pada Masyarakat di Kota/Kabupaten di Provinsi Jawa Timur Tahun 2018," vol. 5, no. 5, pp. 623–632, 2023.

[10] H.Alrasyid, A.Homaidi, M.Kom, Z.Fatah, andM.Kom, "Comparison Support Vector Machine and Random Forest Algorithms in Detect Diabetes," vol. 1, no. 1, pp. 447–453, 2024.

[11] I.Saha, J. P.Sarkar, andU.Maulik, "Ensemble-based rough fuzzy clustering for categorical data," *Knowledge-Based Syst.*, vol. 77, pp. 114–127, 2015, doi: 10.1016/j.knosys.2015.01.008.

[12] A.Jalil, A. Homaidi, and Z.Fatah, "Implementasi Algoritma Support Vector Machine Untuk Klasifikasi Status Stunting Pada Balita," *G-Tech J. Teknol. Terap.*, vol. 8, no. 3, pp. 2070–2079, 2024, doi: 10.33379/gtech.v8i3.4811.

[13] I.Amal, E. W.Pamungkas, S.Kom, M.Kom, andD.Ph, "Aplikasi Pendeteksi Berita Palsu Bahasa Indonesia Menggunakan Framework Flask Dan Streamlit Serta Algoritma Machine Learning Teknik Informatika , Fakultas Komunikasi dan Informatika , Universitas Muhammadiyah Surakarta Kata Kunci : algoritma machine learn," pp. 1–18.

[14] H.Apriyani, "Perbandingan Metode Naïve Bayes Dan Support Vector Machine Dalam

Klasifikasi Penyakit Diabetes Melitus," vol. 1, no. 3, pp. 133–143, 2020.
[15] D. Trisnawati, M. Windarti, I. Sulistyowati, and F. B. Hartono, "Sistem Pakar Pendiagnosa Penyakit Diabetes," vol. 2, no. 1, pp. 14–19, 2022, doi: 10.54840/jcstech.v2i1.32.
[16] B. R. Asmoro, A. Wibowo, and A. F. Aryadi, "Penggunaan Algoritma K-Means Untuk Menganalisa Penjualan Di Bigmart," *Jmari*, vol. 3, no. 2, pp. 189–199, 2022, doi: 10.33050/jmari.v3i2.2427.
[17] U.Kusnia and F.Kurniawan, "Analisis Sentimen Review Aplikasi Media Berita Online Pada Google Play menggunakan Metode Algoritma Support Vector Machines (SVM) Dan Naive Bayes," vol. 5, no. 36, pp. 24–28, 2022.
[18] F.Zafira, B.Irawan, andA.Bahtiar, "Penerapan Data Mining Untuk Estimasi Stok Barang Dengan Metode K-Means Clustering," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 156–161, 2024, doi: 10.36040/jati.v8i1.8319.
[19] H. S. D.Suparwito, R. (Universitas S. D.Gunawan, I. (Universitas S. D.Binanto, R. (Universitas S. D.Arum Kumalasanti, andW. (Universitas S. D.Widyastuti, *Pengantar Pembelajaran Mesin Menggunakan Bahasa Pemrograman Python*. 2023.
[20] R. A.Muhadi, *Pengaruh Hambatan Samping Terhadap Kinerja Jalan (Studi Kasus Ruas Jalan Pekiringan Kecamatan Pekalipan Kota Cirebon)*. repository.unpas.ac.id, 2022. [Online]. Available: http://repository.unpas.ac.id/60972/
[21] A.Ali, "Sistem Informasi Model Prediksi Penyakit Diabetes Menggunakan Algoritma Random Forest di Fasyankes".
[22] A.Setiawan, Z. H.Nst, Z.Khairi, andL.Efrizoni, "KLASIFIKASI TINGKAT RISIKO DIABETES MENGGUNAKAN ALGORITMA," vol. 7, no. 2, pp. 263–271, 2024.
[23] T. Tukino and F.Fifi, "Penerapan Support Vector Machine Untuk Analisis Sentimen Pada Layanan Ojek Online," *J. Desain Dan Anal. Teknol.*, vol. 3, no. 2, pp. 104–113, 2024, doi: 10.58520/jddat.v3i2.59.
[24] E.Tohidi, R.Perdana Herdiansyah, E.Wahyudin, andK.Kaslani, "Analisa Sentimen Komentar Video Youtube Di Channel Tvonenews Tentang Calon Presiden Prabowo Subianto Menggunakan Support Vector Machine," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 660–667, 2024, doi: 10.36040/jati.v8i1.8560.
[25] F. Sains and U.Muhammadiyah, "Sistem Prediksi Penyakit Jantung Berbasis Web Menggunakan Metode SVM dan Framework Streamlit," vol. 4, no. 2, pp. 442–452, 2023.
[26] Z. S.R Azhari, "Hubungan Aktivitas Fisik Dengan Kadar Glukosa Darah Pada Penyandang Diabetes Melitus Tipe 2 Di Wilayah Perumahan Bugel Mas Indah RW 009," vol. 2, no. 7, pp. 86–90, 2022.