

Aplikasi Berbasis Web Berdasarkan Model Klasifikasi Algoritma SVM dan *Logistic Regression* Terhadap Data Diabetes

¹Nita Fitriyani, ²Dinda Resna Amalia, ³Hanny Hikmayanti Handayani, ⁴Anis Fitri Nur Masruriyah
^{1,2,3,4}Universitas Buana Perjuangan Karawang
Karawang, Indonesia

¹if20.nitafitriyani@mhs.ubpkarawang.ac.id, ²if20.dindaamalia@mhs.ubpkarawang.ac.id,
³hanny.hikmayanti@ubpkarawang.ac.id, ⁴anis.masruriyah@ubpkarawang.ac.id

*Penulis Korespondensi

Diajukan : 12/09/2023

Diterima : 26/09/2023

Dipublikasi : 02/10/2023

ABSTRAK

Berdasarkan *International Diabetes Federation Atlas Tenth* edisi 2021, jumlah penderita diabetes mencapai 537 juta orang dalam rentang usia 20-79 tahun. Jumlah penderita diabetes akan terus meningkat mencapai 643 juta pada tahun 2030, bahkan diperkirakan mencapai 783 juta pada tahun 2045. Diabetes tidak hanya menjadi penyebab 6,7 juta kematian, tetapi juga menguras dana kesehatan hingga 966 miliar USD. Tingkat kadar gula darah yang tinggi secara kronis menjadi tanda penyakit diabetes, keadaan ini terjadi ketika tubuh tidak mampu menghasilkan insulin secara efektif. Penelitian ini bertujuan untuk mengembangkan model klasifikasi penderita penyakit diabetes dengan membandingkan dua Algoritma, *Support Vector Machine* (SVM) dan Regresi Logistik. Dalam penelitian ini, model dievaluasi menggunakan metode K-Fold cross validation dengan membagi dataset menjadi 10 subset. Salah satu subset dipilih sebagai data uji, sementara subset lainnya digunakan sebagai data latih. Hasil penelitian menunjukkan bahwa klasifikasi terbaik diperoleh pada Algoritma SVM dengan teknik SMOTE. Model ini mencapai rata-rata *accuracy* sebesar 88,77%, *precision* 88,50%, dan *recall* 89,21%. Dengan demikian, model yang dikembangkan menggunakan Algoritma SVM dengan SMOTE dapat diimplementasikan ke dalam sebuah sistem klasifikasi penyakit diabetes. Pembuatan aplikasi ini ditujukan kepada pihak medis untuk membantu dalam menguatkan diagnosa pemeriksaan, apakah seseorang menderita penyakit diabetes atau tidak dengan tingkat akurasi yang baik.

Kata Kunci: Diabetes, Klasifikasi, Regresi Logistik, *Support Vector Machine*

I. PENDAHULUAN

Diabetes adalah salah satu penyakit mematikan, disebabkan karena tidak terkontrolnya kadar gula darah didalam tubuh, sehingga pankreas tidak dapat menghasilkan insulin dengan cukup (Monikasari et al., 2023). Jika tidak diobati, akan banyak komplikasi penyakit lain yang timbul. Menurut International Diabetes Federation (IDF) Atlas Tenth edisi 2021, terdapat fakta, angka, dan proyeksi terkini mengenai diabetes seluruh dunia. Pada tahun 2021, sekitar 537 juta orang berusia 20-79 tahun menderita diabetes. Jumlah penderita diabetes meningkat menjadi 643 juta pada tahun 2030 dan diperkirakan mencapai 783 juta pada tahun 2045. Diabetes menyebabkan sekitar 6,7 juta kematian dan mencapai 966 miliar USD dalam pengeluaran kesehatan (Federation, 2021). Berdasarkan Pusat Data dan Informasi Kementerian Kesehatan Republik Indonesia, hasil Riskesdas 2018 menurut diagnosis dokter pada usia lebih dari sama dengan 15 tahun prevalensi diabetes di Indonesia sebesar 2%, sedangkan dalam pemeriksaan gula darah sebesar 8,5% (Kementerian Kesehatan RI., 2020).

Salah satu solusi untuk masalah diabetes dalam bidang ilmu komputer adalah pemodelan

untuk melakukan klasifikasi (Karo Karo & Hendriyana, 2022)(Trianto et al., 2022)(Pahlevi et al., 2021)(Kurniadi et al., 2021). Berdasarkan penelitian sebelumnya, Febyan Sulastris dkk menggunakan Algoritma Support Vector Machine (SVM) dalam melakukan komparasi akurasi dengan mengoptimalkan setiap kernel. Kernel rbfdot dan baseldot mendapatkan nilai akurasi yang sama tertinggi dari kernel lain dengan nilai 72.73% (Sulastris et al., 2023). Penelitian mendeteksi penyakit diabetes dengan Algoritma Artificial Neural Network (ANN) dan menggunakan metode CRISP-DM. Data pada penelitian tersebut menghilangkan missing value serta outlier, setelah itu men-splitting data yang kemudian dibagi menjadi data training, testing, dan validasi. Akurasi terbaik mencapai 88% (K et al., 2022). Kajian prediksi diabetes dengan Algoritma Naive Bayes menggunakan operator normalize pada tool rapidminer. Model mampu memprediksi class recall true negative 93.63% dan false negative 97.99%. Nilai precision mencapai 97,64% dan false negative 94,54% dari 1035 diagnosa diabetes. Nilai akurasi 95,94% dengan hasil prediksi 487 true positif dan 548 true negatif (Supandi et al., 2022). Penelitian menggunakan metode Z-Score dengan Algoritma Decision Tree, Naive Bayes, serta SVM. Kombinasi Algoritma SVM dan Z-Score memberi kinerja terbaik dengan *accuracy* 80,73%, dan *f1* 76% (Karo Karo & Hendriyana, 2022). Penelitian Algoritma SVM dan forward selection menghasilkan nilai model SVM dengan akurasi 91,2%, presisi 93,0%, recall 94,3%, dan skor *f1* 93,7% (Wafa et al., 2022).

Dalam penelitian lainnya, Trianto dkk menggunakan Algoritma SVM dalam klasifikasi penyakit diabetes. Untuk mengatasi data yang tidak seimbang, penelitian ini menggunakan Borderline-SMOTE, serta Grid Search dan Bagging. Hasil menunjukkan akurasi 92,1%, presisi 95,51% sehat dan 86,12% diabetes, recall 92,32% sehat dan 91,66% diabetes, *f1-score* 93,39% sehat dan 88,81% diabetes (Trianto et al., 2022). Penelitian menggunakan Algoritma ID3 dan SVM dengan hasil algoritma paling efektif yaitu SVM linear, mencapai akurasi sebesar 78,5%, presisi 79%, recall 56%, dan *f1-score* 66% (Pahlevi et al., 2021). Kajian perbandingan Algoritma KNN dan Logistic Regression pada klasifikasi diabetes. Hasil akurasi pada Algoritma KNN mencapai 85,06%, sedangkan Algoritma Regresi Logistik sebesar 77,92% (Kurniadi et al., 2021). Penelitian menggunakan Algoritma Regresi Logistik. Hasil evaluasi memakai normalisasi dengan akurasi 76%, presisi 66%, dan recall 55%. Sedangkan evaluasi tanpa normalisasi mendapat akurasi 75%, presisi 83%, recall 43% (Cahyani et al., 2022). Penelitian menggunakan Algoritma Logistic Regression. Hasil implementasi Algoritma Logistic Regression menggunakan liblinear dengan normalisasi memiliki nilai recall sebesar 58% (Wahyudi & Kadyanan, 2023).

Berdasarkan paparan di atas, penelitian sebelumnya tidak menggunakan teknik smote pada modeling dan k-fold, sehingga dapat menyebabkan hasil evaluasi yang kurang stabil dan estimasi yang tidak akurat pada kinerja model. Adapun hasil dari model penelitian sebelumnya juga tidak diimplementasikan ke dalam sebuah sistem. Sedangkan, mendiagnosis apakah seseorang terkena penyakit diabetes atau tidak menjadi tujuan penelitian ini. Untuk mencapai tujuan tersebut, maka penelitian ini menerapkan perbandingan algoritma, yakni Regresi Logistik dan *Support Vector Machine* (SVM) dalam pembuatan modelnya, dengan menerapkan teknik SMOTE dan evaluasi menggunakan K-fold cross validation. Setelah itu, model yang paling baik akan diimplementasikan ke dalam sebuah sistem yang dapat mengklasifikasi secara benar.

II. STUDI LITERATUR

Diabetes

Diabetes merupakan suatu keadaan kedokteran yang ditandai oleh tingginya kandungan gula darah didalam tubuh secara kronis. Keadaan ini terjadi saat tubuh tidak mampu menggunakan ataupun menghasilkan insulin secara optimal. Insulin merupakan hormon yang diperlukan untuk mengendalikan kandungan gula darah (Nurani & Fitriyani, 2023). Gejala umum pada penyakit diabetes mencakup rasa haus yang berlebih, sering buang air kecil, penurunan berat badan, kelelahan, dan luka yang sulit sembuh (Aldi, 2022). Diabetes memerlukan perawatan yang teratur, termasuk pengaturan pola makan, aktivitas fisik, dan pemantauan rutin kadar gula darah (Munadia et al., 2023).

Algoritma Support Vector Machine



Konsep dasar metode ini melibatkan pencarian ruang pemisah yang optimal, disebut hiperplane. Dalam kehidupan sehari-hari, sering terjadi masalah yang kompleks atau data tidak bisa dipisahkan secara linear, sehingga tidak ada garis atau bidang yang berfungsi sebagai pemisah antara kelas data tersebut (Maksur, 2020). Untuk mengatasi permasalahan ini, Algoritma SVM menggunakan fungsi kernel untuk memetakan data ke dalam ukuran yang lebih tinggi. Terdapat empat fungsi kernel umum yang digunakan SVM, seperti kernel linier, kernel polinomial, kernel Fungsi Basis Radial (RBF), dan kernel sigmoid (Yuliati et al., 2020).

Algoritma Regresi Logistik

Regresi Logistik merupakan metode klasifikasi untuk memahami hubungan antara satu variabel respons kategorikal dengan beberapa variabel prediktor yang bersifat kontinu. Tujuan utama dari Regresi Logistik adalah untuk memprediksi probabilitas atau kemungkinan terjadinya suatu kejadian berdasarkan variabel prediktor yang terkait. Dalam proses Regresi Logistik, data digunakan untuk melatih (fit) model, kemudian model tersebut dievaluasi menggunakan data pengujian yang tidak digunakan dalam proses pelatihan (Wardhani et al., 2019).

Confusion Matrix

Confusion matrix adalah metode evaluasi yang digunakan dalam analisis hasil klasifikasi pada machine learning. Terdapat empat istilah dalam confusion matrix, yaitu True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). TP mewakili data positif yang terdeteksi dengan benar, FN mewakili data yang sebenarnya positif tetapi salah terdeteksi sebagai negatif. TN mewakili data negatif dan terdeteksi benar, FP mewakili data yang sebenarnya negatif tetapi salah terdeteksi sebagai positif (Tjahjadi & Santoso, 2023). Dengan nilai-nilai tersebut, metrik evaluasi seperti akurasi, presisi, dan recal dapat dihitung untuk mengevaluasi performa keseluruhan dari model.

K-Fold Cross Validation

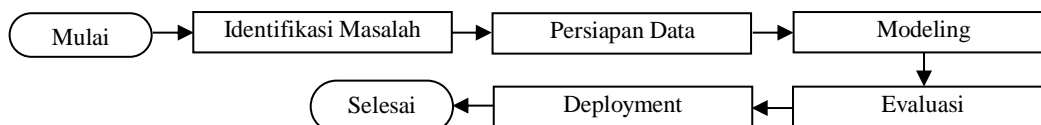
K-Fold cross validation adalah metode evaluasi yang umum digunakan untuk mengukur kinerja model secara objektif. Metode ini melibatkan pemisahan data menjadi k subset atau "fold" dengan ukuran yang sama. Dalam setiap iterasi, salah satu subset digunakan sebagai data uji, sementara subset lainnya sebagai data latih. Proses tersebut diulang sebanyak k kali, di mana setiap subset akan menjadi data uji satu kali. Hasil evaluasi dari setiap iterasi dijumlahkan dan dihitung rata-ratanya untuk mendapatkan matrix kinerja yang lebih akurat (Amalia, 2023).

Rapid Application Development

Analisis kebutuhan adalah tahapan pertama dalam metode Software Development Life Cycle (SDLC). Tahapan ini merupakan proses untuk memahami dan mendokumentasikan kebutuhan pengguna dalam sebuah sistem yang akan dibangun (Purnasari et al., 2022). Rapid Application Development (RAD) merupakan suatu metode pengembangan software yang siklusnya relatif singkat, dengan menekankan siklus hidup pengembangan pendek dan mewakili tipe adaptasi cepat dari metode Waterfall dengan konstruksi komponen (Hariyanto et al., 2021).

III. METODE

Tahapan kerja penelitian ini mencakup identifikasi masalah, persiapan data, modeling, evaluasi, dan deployment, disertai beberapa visualisasi data.



Gambar 1. Tahapan Penelitian

Identifikasi Masalah

Diabetes merupakan masalah serius yang semakin meningkat prevalensinya di masyarakat.

Kurangnya pengetahuan dan kesadaran tentang diabetes menjadi kendala dalam upaya pencegahan penyakit diabetes. Oleh karena itu, penelitian dan inovasi pengobatan diabetes sangat penting dalam meningkatkan pendekatan pemeriksaan dengan menemukan solusi yang efektif, seperti membuat sistem untuk mendiagnosa seseorang terkena penyakit diabetes atau tidak.

Persiapan Data

Persiapan data dilakukan untuk memastikan data yang akan digunakan dalam penelitian siap untuk analisis lebih lanjut. Tahapan persiapan data mencakup beberapa langkah, seperti pembersihan data dan pemilihan fitur yang relevan. Dalam pembersihan data, langkah-langkahnya seperti menghapus nilai yang hilang atau outlier, mengisi nilai yang hilang, dan mengatasi duplikasi data dilakukan. Seluruh langkah persiapan data ini bertujuan untuk memastikan kualitas dan kebersihan data sebelum masuk ke tahap analisis dan pemodelan.

Modeling

Modeling dilakukan sebelum tahap implementasi deployment. Proses modeling ini menerapkan algoritma untuk menghasilkan model yang baik dan sesuai dengan kebutuhan. Model dibangun menggunakan Algoritma Regresi Logistik dan *Support Vector Machine* (SVM) dengan teknik SMOTE.

Evaluasi

Dalam Penelitian ini, Evaluasi model menggunakan K-Fold cross validation untuk menghasilkan *accuracy*, *precision*, dan *recall* pada penelitian ini. Evaluasi K-Fold digunakan untuk menguji dan mengukur kinerja kedua algoritma tersebut secara objektif.

Deployment

Tahap deployment menghasilkan sebuah sistem berbasis website yang memungkinkan pengguna untuk menguji data klasifikasi penyakit diabetes. Sistem ini dibuat menggunakan streamlit yang merupakan sebuah framework dengan bahasa pemrograman python dalam membangun aplikasi web yang responsif, serta dapat diakses melalui browser. Sistem dirancang berdasarkan hasil proses modeling yang telah dipilih sebelumnya dengan performa terbaik.

IV. HASIL DAN PEMBAHASAN

Persiapan Data

Dataset diambil dari National Institute of Diabetes and Digestive and Kidney Diseases yang disediakan melalui Kaggle. Tujuan dari proses persiapan data ini adalah untuk melakukan klasifikasi diagnosa apakah seseorang terkena penyakit diabetes atau tidak. Dataset terdiri dari 8 variabel prediktor medis dan satu variabel target (outcome), dengan total sampel 768 baris. Atribut dalam dataset ini dikumpulkan dari wanita keturunan India Pima dengan usia minimum 21 tahun.

Tabel 1. Dataset

Preg	Glu	Blo	Skin	Ins	BMI	Ped	Age	Class
1	85	66	29	0	26.6	0.351	31	0
8	183	64	0	0	23.3	0.672	32	1
0	137	40	35	168	43.1	2.288	33	1

Sumber: (Cahyani & Basuki, 2019)

Tabel 2. Deskripsi Atribut

Atribut	Deskripsi
Pregnancies	Jumlah kehamilan
Glucose	Konsentrasi glukosa setelah 2 jam tes
BloodPressure	Tekanan darah diastolik
SkinThickness	Ketebalan kulit bagian triceps
Insulin	Konsentrasi insulin setelah 2 jam

BMI	Indeks massa tubuh (berat kg / tinggi m) ²
Pedigree	Riwayat diabetes dalam keluarga
Age	Umur
Class	Klasifikasi. 1 berarti terkena, 0 berarti tidak

Sumber: (Cahyani & Basuki, 2019)

Preprocessing dilakukan dengan tujuan untuk mengidentifikasi dan menangani data yang hilang (missing value), karena hal tersebut mengakibatkan masalah dalam proses pelatihan model. Missing value dapat terlihat dari nilai 0 dalam atribut yang seharusnya tidak boleh memiliki nilai 0, seperti glucose, bloodPressure, skinThickness, insulin, BMI, pedigree, dan Age. Pada proses ini ditemukan lima atribut yang memiliki nilai 0, yaitu Glucose, BloodPressure, SkinThickness, Insulin, serta BMI. Sehingga solusi yang dilakukan adalah dengan mengubah nilai 0 pada atribut tersebut menjadi NaN, kemudian diganti menjadi nilai yang sesuai berdasarkan nilai tengah (median) dari variabel outcome.

Tabel 3. Nilai Rata-rata

Outcome	Glucose	BloodPres	SkinThick	Insulin	BMI
0	107.0	70.0	27.0	102.5	30.1
1	140.0	74.5	32.0	169.5	34.3

Sumber: Hasil pengolahan data

Setelah mengisi nilai yang hilang, langkah selanjutnya adalah menghilangkan noise pada data. Noise dihilangkan menggunakan metode Z-score untuk mengeliminasi data yang tidak valid atau tidak konsisten. Proses penghilangan noise membuat dataset menjadi lebih bersih, sehingga analisis dapat dilakukan dengan lebih akurat. Setelah proses penghilangan noise, jumlah kolom pada dataset menjadi 721.

Langkah selanjutnya adalah mendeteksi dan menangani outlier pada data. Hal ini dilakukan menggunakan metode Inter Quartile Range (IQR) atau melalui analisis visual menggunakan boxplot. Outlier adalah nilai-nilai yang secara signifikan berbeda dari nilai lain dalam dataset dan dapat mempengaruhi hasil analisis statistik. Setelah dilakukan pembersihan data yang mengandung outlier, jumlah kolom pada dataset menjadi 572.

Setelah itu, dilakukan proses normalisasi data menggunakan Min-Max Scaler untuk mengubah rentang nilai variabel-variabel dataset ke dalam rentang standar yang ditentukan, antara 0 hingga 1 atau -1 hingga 1. Dengan normalisasi, maka dapat memperlakukan semua variabel dalam proporsi yang sama, sehingga meminimalkan bias yang mungkin timbul akibat perbedaan skala antar variabel.

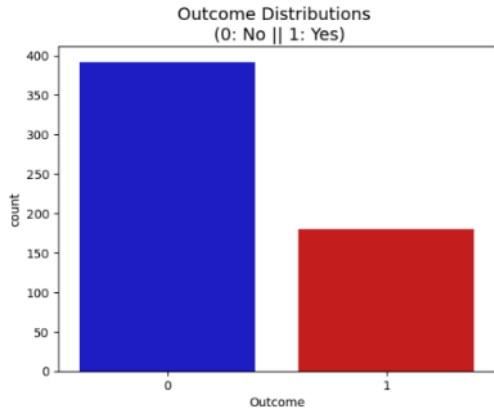
Tabel 4. Nilai Normalisasi

Pregna	Glucu	Blood	Skin	Insulin	BMI	Ped	Age
0.461538	0.684211	0.466667	0.740741	0.604331	0.518519	0.529923	0.674419
0.076923	0.269737	0.366667	0.518519	0.340551	0.282828	0.263514	0.232558

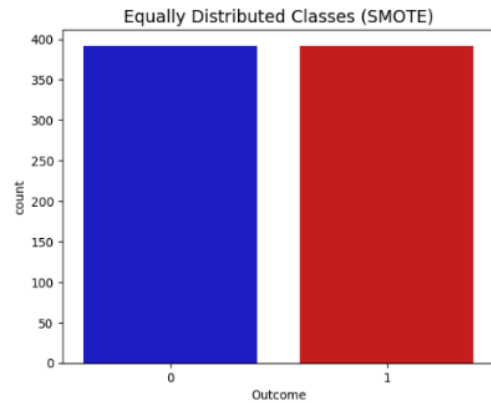
Sumber: Hasil pengolahan data

Modeling

Selanjutnya, dilakukan modeling menggunakan Algoritma Regresi Logistik dan SVM jenis kernel linear dengan Synthetic Minority Over-sampling Technique (SMOTE). SMOTE adalah metode yang digunakan untuk menghasilkan sampel sintesis dari kelas minoritas sehingga menciptakan keseimbangan antara kelas mayoritas dan minoritas. Dengan menerapkan SMOTE, maka dapat meningkatkan representasi kelas minoritas dalam dataset, yang meningkatkan performa model dan evaluasi selanjutnya.



Gambar 2. Dataset sebelum SMOTE



Gambar 3. Dataset sesudah SMOTE

Sumber: Hasil pengolahan data

Evaluasi

Proses evaluasi dimulai dengan memisahkan data menjadi k subset dalam ukuran yang sama, di mana k yang dipilih adalah 10. Kemudian setiap subset secara bergantian digunakan sebagai training set, sementara subset lainnya untuk testing set. Proses ini melibatkan pelatihan model Regresi Logistik dan SVM jenis kernel linear pada setiap iterasi K-Fold. Setelah proses pelatihan model selesai, dilakukan penghitungan matriks evaluasi seperti nilai *accuracy*, *precision*, dan *recall* untuk mengukur performa model. Evaluasi ini memberikan informasi sejauh mana model mampu melakukan klasifikasi secara tepat, mengidentifikasi kasus positif dengan presisi baik, dan mengenali seluruh kasus positif dengan tingkat recall tinggi.

Tabel 5. Performa Regresi Logistik

Nilai	K-Fold									
	1	2	3	4	5	6	7	8	9	10
Tanpa SMOTE										
<i>Accuracy</i>	0.82	0.89	0.87	0.80	0.80	0.85	0.84	0.91	0.85	0.80
<i>Precision</i>	0.77	0.88	0.76	0.83	0.50	1.00	0.86	0.81	0.90	0.73
<i>Recall</i>	0.70	0.61	0.88	0.52	0.54	0.61	0.65	0.86	0.76	0.61
Dengan SMOTE										
<i>Accuracy</i>	0.84	0.88	0.89	0.87	0.83	0.84	0.88	0.89	0.91	0.79
<i>Precision</i>	0.83	0.85	0.83	0.84	0.85	0.80	0.87	0.92	0.92	0.75
<i>Recall</i>	0.83	0.92	0.97	0.89	0.87	0.92	0.89	0.88	0.90	0.79

Sumber: Hasil pengolahan data

Tabel 6. Performa SVM Kernel Linear

Nilai	K-Fold									
	1	2	3	4	5	6	7	8	9	10
Tanpa SMOTE										
<i>Accuracy</i>	0.79	0.93	0.84	0.85	0.85	0.92	0.85	0.91	0.89	0.85
<i>Precision</i>	0.68	0.80	0.68	0.78	0.60	1.00	0.83	0.75	0.91	0.77
<i>Recall</i>	0.75	0.92	0.94	0.78	0.81	0.80	0.75	1.00	0.84	0.77
Dengan SMOTE										
<i>Accuracy</i>	0.88	0.91	0.91	0.87	0.87	0.87	0.88	0.88	0.91	0.87
<i>Precision</i>	0.88	0.89	0.85	0.82	0.87	0.83	0.89	0.92	0.95	0.90
<i>Recall</i>	0.86	0.92	0.97	0.91	0.91	0.92	0.86	0.86	0.88	0.79

Sumber: Hasil pengolahan data

Setelah melakukan evaluasi pada setiap lipatan dalam 10-Fold, hasil evaluasi tersebut dihitung dengan mengambil rata-rata dari seluruh nilai *accuracy*, *precision*, dan *recall* yang diperoleh pada masing-masing model. Proses ini bertujuan untuk menggambarkan kinerja model secara keseluruhan berdasarkan performa yang diamati.

Tabel 7. Perbandingan Model Regresi Logistik dan SVM

Metode	Accuracy	Precision	Recall
LR	84.96%	80.79%	67.82%
LR+SMOTE	86.72%	85.02%	88.97%
SVM	87.41%	78.25%	84.02%
SVM+SMOTE	88.77%	88.50%	89.21%

Sumber: Hasil pengolahan data

Dari hasil perbandingan evaluasi model tanpa SMOTE pada penelitian ini menghasilkan nilai tertinggi pada Algoritma SVM kernel linear dengan *accuracy* 87,41% dan *recall* 84,02%, sedangkan nilai *precision* tertinggi 80.79% terdapat pada Algoritma Regresi Logistik. Adapun hasil evaluasi model dengan SMOTE tertinggi pada Algoritma SVM kernel linear dengan *accuracy* 88,77%, *precision* 88,50%, dan *recall* 89,21%. Hasil tersebut menunjukkan bahwa penggunaan SMOTE pada dataset meningkatkan nilai *accuracy*, *precision*, dan *recall* yang dapat mengindikasikan kemampuan model dalam mengenali kelas diabetes. Selanjutnya, hasil dari model diimplementasikan pada sistem menggunakan streamlit.

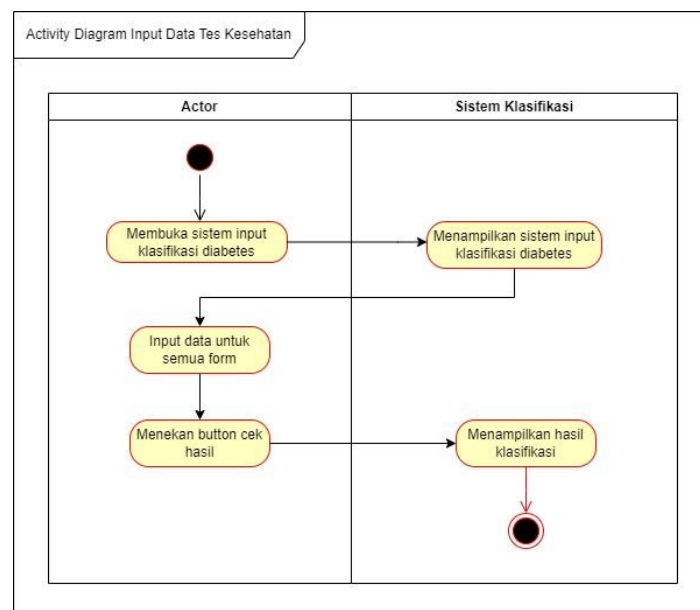
Deployment

Model yang telah dibuat diimplementasikan ke dalam sebuah sistem. Sistem dibangun menggunakan metode RAD dengan empat tahapan.

a. Perencanaan Kebutuhan

Pembuatan aplikasi ini ditujukan kepada pihak medis untuk membantu dalam menguatkan diagnosa apakah seseorang menderita diabetes atau tidak. Pengujian dilakukan dengan menginputkan nilai dari setiap atribut seperti pregnancies, glucose, age, dan lainnya ke dalam sistem. Hasil diagnosa akan ditampilkan kepada pengguna melalui antarmuka website, sehingga pihak medis dapat memberikan informasi tentang status kesehatan seseorang berdasarkan data yang diinputkan.

b. Desain Sistem



Gambar 4. Activity diagram sistem klasifikasi diabetes

Pada gambar 4, hal pertama yang dilakukan oleh tenaga medis yaitu membuka sistem klasifikasi diabetes, kemudian sistem akan menampilkan halaman untuk menginputkan data tes kesehatan. Lalu tenaga medis mengisi semua kolom dan menekan tombol cek hasil jika sudah. Sistem akan menampilkan hasil klasifikasi dari semua data yang diinputkan.

c. Prototype

Hal pertama yang dilakukan oleh tenaga medis yaitu membuka sistem klasifikasi diabetes, kemudian sistem akan menampilkan halaman untuk menginputkan data tes kesehatan. Lalu tenaga medis mengisi semua kolom dan menekan tombol cek hasil jika sudah. Sistem akan menampilkan hasil klasifikasi dari semua data yang diinputkan.

The image shows a web browser window with the address bar containing 'https://klasifikasi-diabetes'. The main content area is titled 'Klasifikasi Diabetes' and contains a form with eight input fields. The fields are organized into two columns. The left column has four fields: 'Nilai Pregnancies', 'Nilai Glucose', 'Nilai Blood Pressure', and 'Nilai Skin Thickness'. The right column has four fields: 'Nilai Insulin', 'Nilai BMI', 'Nilai Diabetes Pedigree Function', and 'Nilai Age'. At the bottom left of the form area, there is a button labeled 'Cek Hasil'.

Gambar 5. Mockup sistem klasifikasi diabetes

Pada gambar 5 mockup sistem klasifikasi diabetes memiliki delapan kolom yang harus diisi oleh tenaga medis untuk mengetahui hasil tes kesehatan pasien. Kolom terdiri dari nilai pregnancies, glucose, blood pressure, skinthickness, insulin, BMI, pedigree function, dan age.

d. Implementasi

Hasil tahapan perencanaan kebutuhan, desain sistem, dan prototype diimplementasikan menjadi sebuah sistem. Sistem ini menjadi alat bagi tenaga medis dalam memahami dan mengelola kasus diabetes, mempercepat proses diagnosa, dan memberikan pengobatan yang tepat sesuai dengan kondisi pasien.

Pada gambar 6 dibawah ini, hasil implementasi model ke dalam sistem untuk mengetahui kondisi kesehatan pasien diabetes. Sistem tersebut hanya dapat digunakan oleh tenaga medis dengan mengisi semua kolom berdasarkan hasil tes kesehatan. Kolom yang harus diisi oleh tenaga medis yaitu nilai pregnancies, glucose, blood pressure, skinthickness, insulin, BMI, pedigree function, dan age. Setelah tenaga medis mengisi semua kolom, sistem akan menampilkan hasil dari data yang diinputkan, apakah seorang pasien terkena diabetes atau tidak.

Klasifikasi Diabetes

Nilai Pregnancies	Nilai Insulin
3	245
Nilai Glucose	Nilai BMI
158	31.6
Nilai Blood Pressure	Nilai Diabetes Pedigree Function
76	0.851
Nilai Skin Thickness	Nilai Age
36	28

Pasien terkena Diabetes

Gambar 6. Hasil Pengujian Sistem

V. KESIMPULAN

Dapat dilihat pada Algoritma SVM maupun Regresi Logistik, perbedaan terlihat pada nilai *accuracy*, *precision*, dan *recall* antara dataset dengan SMOTE dan tanpa SMOTE. Berdasarkan hasil pengujian, model klasifikasi Algoritma SVM dengan SMOTE memiliki hasil terbaik dibandingkan model lainnya. Dataset pada Algoritma SVM tanpa SMOTE memiliki *accuracy* 87,41%, *precision* 78,25%, dan *recall* 84,02%. Sedangkan dataset Algoritma SVM dengan SMOTE memiliki *accuracy* 88,77%, *precision* 88,50%, dan *recall* 89,21%. Selanjutnya, dataset pada Algoritma Regresi Logistik tanpa SMOTE diperoleh nilai *accuracy* 84,96%, *precision* 80,79%, dan *recall* 67,82%. Sementara Regresi Logistik dengan SMOTE diperoleh nilai *accuracy* 86,72%, *precision* 85,02%, dan *recall* 88,97%. Hasil tersebut menunjukkan bahwa penggunaan SMOTE pada dataset dapat meningkatkan nilai *accuracy*, *precision*, dan *recall* yang mengindikasikan kemampuan model dalam mengenali kelas diabetes.

VI. REFERENSI

- Aldi, M. (2022). *Studi Kasus Asuhan Keperawatan Pada Klien Yang Mengalami Diabetes Melitus Tipe II Di Wilayah Desa Sedulang Kecamatan Muara Kaman*.
- Amalia, N. L. (2023). *Analisis Sentimen Pada Perpindahan Ibukota Indonesia Dengan Algoritma Support Vector Machine: Evaluasi Lesikon, Metode Ekstraksi Fitur, Dan Kernel Trick*.
- Cahyani, Q. R., Finandi, M. J., Rianti, J., Arianti, D. L., & Putra, A. D. P. (2022). Prediksi Risiko Penyakit Diabetes menggunakan Algoritma Regresi Logistik. *JOMLAI: Journal of Machine Learning and Artificial Intelligence*, 1(2), 107–114. <https://doi.org/10.55123/jomlai.v1i2.598>
- Federation, I. D. (2021). *Diabetes facts & figures*. International Diabetes Federation.
- Hariyanto, D., Sastra, R., & Putri, F. E. (2021). Implementasi Metode Rapid Application Development Pada Sistem Informasi Perpustakaan. *Jurnal JUPITER*, 13(1), 110–117.
- K, C. M., Haryatmi, E., Fajriatifah, R. Y., & Puspita, Y. H. (2022). Penyakit Diabetes menggunakan Algoritma Artificial Neural Network. *JURNAL DATA SCIENCE & INFORMATIKA*, 2(1), 46–52.

- Karo Karo, I. M., & Hendriyana. (2022). Klasifikasi Penderita Diabetes Menggunakan Algoritma Machine Learning Dan Z-Score. *Jurnal Teknologi Terpadu*, 8(2), 94–99.
- Kementerian Kesehatan RI. (2020). Infodatin tetap produktif, cegah, dan atasi Diabetes Melitus 2020. In *Pusat Data dan Informasi Kementerian Kesehatan RI* (pp. 1–10).
- Kurniadi, R. P., Saedudin, R. R., & Widartha, V. P. (2021). Perbandingan Akurasi Algoritma K-Nearest Neighbor Dan Logistic Regression Untuk Klasifikasi Penyakit Diabetes. *E-Proceeding of Engineering*, 8(5), 9757–9764.
- Maksur, I. Al. (2020). *Perancangan Dan Pembuatan Prototype MyBots Pada Media Sosial Discord Serta Klasifikasi SVM*.
- Monikasari, Widyastiti, N. S., Mahati, E., Syauqy, A., & Al-Baarri, A. N. (2023). Pengaruh pemberian ekstrak bekatul beras hitam (*Oryza sativa L. indica*) terhadap kadar MDA, SOD dan trigliserida pada tikus diabetes mellitus tipe 2. *Aceh Nutrition Journal*, 8(1), 129–138.
- Munadia, Syarif, H., & Husna, C. (2023). Indeks Massa Tubuh Dan Kadar Gula Darah Pada Pasien Diabetes Mellitus Tipe 2. *Jurnal Keperawatan Silampari*, 6(2), 1836–1843.
- Nurani, R. D., & Fitriyani. (2023). Progressive Muscle Relaxation (PMR) Untuk Menurunkan Kadar Gula Darah Pada Pasien Diabetes Mellitus Tipe 2. *Jurnal Keperawatan Bunda Delima*, 5(1), 38–45.
- Pahlevi, R., Fredlina, K. Q., & Utami, N. W. (2021). Penerapan Algoritma ID3 Dan SVM Pada Klasifikasi Penyakit Diabetes Melitus Tipe 2. *Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) 2021*, 64–75.
- Purnasari, M., Hartiwi, Y., & Nurhayati, N. (2022). Perancangan Sistem Informasi Pengelolaan Dana Masjid Berbasis Web Menggunakan Unified Modeling Language (UML). *Resolusi : Rekayasa Teknik Informatika Dan Informasi*, 2(6), 258–264.
- Sulastri, F., Sunge, A. S., & Achmad, L. I. (2023). Komparasi Akurasi Kernel Support Vector Machine Dalam Prediksi Diabetes. *Prosiding SAINTEK: Sains Dan Teknologi*, 2(1), 555–562.
- Supandi, A., Faqih, A., & Basysyar, M. (2022). Prediksi Penyakit Diabetes Menggunakan Machine Learning Dengan Algoritma Naïve Bayes. *JURSIMA Jurnal Sistem Informasi Dan Manajemen*, 10(2), 146–152.
- Tjahjadi, E. V., & Santoso, B. (2023). Klasifikasi Malware Menggunakan Teknik Machine Learning. *Jurnal Ilmiah Ilmu Komputer*, 2(1), 60–70.
- Trianto, Muliawati, A., & Irmada, H. N. (2022). Penerapan Borderline-SMOTE dan Grid Search pada Bagging-SVM untuk Klasifikasi Penyakit Diabetes. *Seminar Nasional Mahasiswa Ilmu Komputer Dan Aplikasinya (SENAMIKA)*, 3(2), 102–113.
- Wafa, H. S., Hadiana, A. I., Umbara, F. R., Terusan, J., Sudirman, J., Sel, K. C., & Cimahi, K. (2022). Prediksi Penyakit Diabetes Menggunakan Algoritma Support Vector Machine (SVM). *INFORMATICS AND DIGITAL EXPERT (INDEX)*, 4(1), 40–45.
- Wahyudi, I. W. T., & Kadyanan, I. G. A. G. A. (2023). Implementasi Logistic Regression dalam Sistem Diagnosa Penyakit Diabetes dengan KNN. *Jurnal Elektronik Ilmu Komputer Udayana*, 11(4), 743–750.
- Wardhani, R. P. S., Sudarno, S., & Maruddani, D. A. I. (2019). Optimasi Regresi Logistik Menggunakan Algoritma Genetika Untuk Pemodelan Faktor-faktor Yang Mempengaruhi Penggolongan Kredit Bank (Studi Kasus: Debitur di PT BPR Gunung Lawu Klaten Periode Tahun 2017). *Jurnal Gaussian*, 8(4), 506–517.
- Yuliati, I. F., Wulandary, S., & Sihombing, P. R. (2020). Penerapan Metode Support Vector Machine (SVM) dan Backpropagation Neural Network (BPNN) dalam Pengklasifikasian Pasangan Usia Subur di Jawa Barat. *Jurnal Statistika Dan Aplikasinya*, 4(1), 23–34.