

Analisis Penggunaan Support Vector Machine pada Deteksi Dini Penyakit Diabetes Melitus

Trisya Septiana^{*1}, Mona Arif Muda², Deny Budiyanto³, Mahendra Pratama⁴, Winaldi Putra Jaya⁵

^{1,2,3,4,5}Teknik Informatika, Fakultas Teknik, Universitas Lampung, Indonesia
Email: ¹trisya.septiana@eng.unila.ac.id, ²mona.arifmuda@eng.unila.ac.id,
³deny.budiyanto@eng.unila.ac.id, ⁴mahendra.pratama@eng.unila.ac.id,
⁵winaldi.putra21@students.unila.ac.id

Abstrak

Peningkatan kadar gula dalam darah serta tidak bekerjanya insulin di dalam tubuh merupakan salah satu ciri dari penyakit diabetes melitus. Semakin lama seseorang mengidap penyakit ini, maka dapat menyebabkan komplikasi yang berdampak pada tingginya risiko kematian. Sehingga penyakit ini menjadi salah satu penyakit paling mematikan di dunia. Pengendalian dan pencegahan penyakit ini dapat dilakukan dengan cara melakukan monitoring secara berkala pada seseorang yang memiliki riwayat ataupun pola hidup yang tidak sehat. Salah satu cara monitoring tersebut dengan mendeteksi awal penyakit diabetes melalui identifikasi parameter pendukung dan memanfaatkan algoritma *Machine learning* yaitu Support Vector Machine (SVM). Cara ini dapat menjadi langkah yang signifikan dalam meningkatkan pemahaman dan deteksi dini penyakit tersebut. Penelitian ini menggunakan 9 parameter diantaranya mencakup *pregnancies*, *glucose*, *blood pressure*, *skin thickness*, *insulin*, *body mass index*, *diabetes pedigree function*, *age*, dan *outcome* (sebagai *output*) dari 768 data. SVM dapat memproses data klinis pasien dan membangun model prediktif untuk mengklasifikasikan apakah seseorang mungkin menderita diabetes atau tidak. Sebelum diimplementasikan melalui SVM, dilakukan terlebih dahulu preprocessing data yang terdiri dari Pembersihan *outliers*, pengisian *missing data value*, normalisasi data, penggunaan SMOTE dan pembagian *train* dan *test data*. Dan selanjutnya dirancang analisis model *machine learning* dengan menerapkan *cross validation* dan *grid search validation*. Hasil dari pengolahan data ini didapatkan sebuah sistem yang divisualisasikan dalam sebuah *website* statis yang dapat memudahkan kemampuan diagnosis, pengenalan dini kondisi prediabetes dengan nilai akurasi yaitu 81%.

Kata kunci: Deteksi Dini, Diabetes Melitus, Machine learning, Support Vector Machine

Abstract

The increase in blood sugar levels and the failure of insulin to work in the body are characteristics of diabetes mellitus. The longer a person suffers from this disease, the more it can cause complications that lead to a higher risk of death. Therefore, this disease has become one of the deadliest diseases in the world. Control and prevention of this disease can be done by regularly monitoring individuals who have a history or unhealthy lifestyle. One way to monitor this is by early detection of diabetes through the identification of supporting parameters and utilizing a machine learning algorithm called Support Vector Machine (SVM). This method can be a significant step in improving understanding and early detection of the disease. This study uses 9 parameters including *pregnancies*, *glucose*, *blood pressure*, *skin thickness*, *insulin*, *body mass index*, *diabetes pedigree function*, *age*, and *outcome* (as output) from 768 data. SVM can process clinical patient data and build a predictive model to classify whether a person is likely to have diabetes or not. Before implementing SVM, preprocessing of the data is done first, which includes cleaning outliers, filling missing data values, data normalization, using SMOTE, and splitting train and test data. Furthermore, a machine learning model analysis is designed by applying cross-validation and grid search validation. The result of this data processing is a system visualized in a static website that can facilitate diagnostic capability and early recognition of prediabetes conditions with an accuracy of 81%.

Keywords: Diabetes Mellitus, Early Detection, Machine learning, Support Vector Machine

1. PENDAHULUAN

Ketidakmampuan pankreas untuk mensekresi insulin serta adanya gangguan kerja insulin menyebabkan terjadinya gangguan kerja tubuh yang teridentifikasi melalui hiperglikemia merupakan ciri khas dari penyakit Diabetes. Kondisi hiperglikemik kronis dapat menimbulkan kerusakan berkepanjangan dan tidak berfungsinya beberapa organ, seperti mata, ginjal, jantung, dan pembuluh darah (Tomic et al., 2022). Diabetes, sebagai salah satu masalah kesehatan utama di seluruh dunia, yang ditandai dengan jumlah penderita diabetes yang terus meningkat secara dramatis selama beberapa dekade terakhir. Hal ini juga berdampak pada Indonesia yang berada di antara 10 negara dalam kategori jumlah penderita diabetes terbesar di dunia pada tahun 2019 (Biologi et al., n.d.; Kurniawaty, 2014; Nurmaya Stiani et al., 2022).

Faktor risiko utama seperti obesitas, kurangnya aktivitas fisik, dan keturunan genetik menjadi kontributor utama terhadap peningkatan prevalensi diabetes. Identifikasi dini dan manajemen yang efisien menjadi kunci untuk mengurangi dampak negatif penyakit ini terhadap kesehatan masyarakat. Dengan jumlah kasus diabetes yang semakin meningkat, maka perlu adanya tindakan awal sebagai deteksi dini penyakit diabetes. Untuk itu diperlukan analisis deteksi dini penyakit diabetes melalui implementasi menggunakan algoritma Support Vector Machine (SVM) (Veronica Agustin et al., 2023).

Analisis penyakit diabetes menggunakan SVM adalah langkah signifikan dalam meningkatkan pemahaman dan deteksi dini penyakit tersebut. Dalam upaya meningkatkan metode identifikasi dini, analisis penyakit diabetes menggunakan Support Vector Machine (SVM) menjadi solusi yang menjanjikan. SVM adalah algoritma *machine learning* yang digunakan untuk tugas klasifikasi dan regresi. Dalam konteks diabetes, SVM dapat memproses data klinis pasien dan membangun model prediktif untuk mengklasifikasikan apakah seseorang mungkin menderita diabetes atau tidak. Keunggulan SVM terletak pada kemampuannya mengatasi data yang tidak linear dan memiliki dimensi tinggi. Hal ini sangat relevan dalam analisis diabetes karena data klinis seringkali kompleks dan melibatkan banyak variabel. SVM dapat memberikan hasil yang andal dan dapat diandalkan dalam mengidentifikasi pola dan tren yang mungkin sulit dikenali oleh metode konvensional (Fadli et al., 2021; Gede et al., n.d.).

Melalui penerapan SVM dalam analisis penyakit diabetes, diharapkan dapat meningkatkan kemampuan diagnosis, memungkinkan pengenalan dini kondisi prediabetes, dan memberikan dasar untuk intervensi medis yang tepat waktu (Han et al., 2011; Tan et al., n.d.). Dengan demikian, langkah ini menjadi kontribusi positif dalam upaya membantu dokter dan pasien dalam menentukan apakah seseorang dinyatakan terkena penyakit diabetes atau tidak. Penelitian ini bertujuan untuk mengembangkan dan menerapkan model *machine learning*, khususnya menggunakan algoritma Support Vector Machine (SVM), dalam proses identifikasi penyakit diabetes (Huda Ovirianti et al., 2022; Jun, 2021). Adapun tujuan utama dan spesifik penelitian ini dapat dijabarkan sebagai berikut :

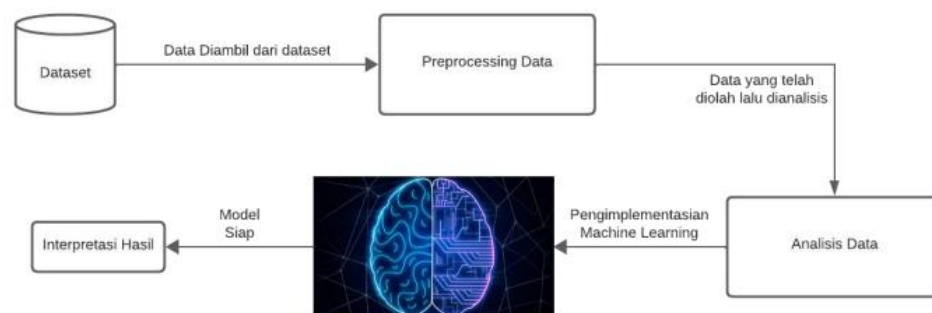
- Meningkatkan Akurasi Identifikasi Penyakit Diabetes. Dengan menerapkan algoritma SVM, diharapkan model dapat memahami pola-pola kompleks dalam data klinis pasien dan menghasilkan prediksi yang akurat terkait dengan keberadaan atau ketiadaan diabetes.
- Mengurangi Waktu Proses Identifikasi. SVM memiliki kemampuan dalam menangani data kompleks, sehingga dapat mempercepat proses diagnosis dan memberikan hasil yang cepat untuk mendukung pengambilan keputusan yang lebih efisien.
- Integrasi Model dengan Sistem Kesehatan. Seiring dengan kemajuan teknologi, tujuan ini mencakup integrasi model identifikasi penyakit diabetes menggunakan SVM dengan sistem kesehatan yang ada. Integrasi ini memfasilitasi penggunaan model dalam lingkungan klinis sehari-hari.
- Membantu Profesional Kesehatan dalam Pengambilan Keputusan: *Machine learning* yang sudah diintegrasikan dapat membantu profesional kesehatan, seperti dokter, dalam mempertimbangkan kemungkinan seseorang menderita diabetes. Dengan adanya penerapan ini, diharapkan profesional kesehatan dapat mengambil keputusan dengan lebih baik, mempercepat intervensi, dan meningkatkan manajemen pasien.

2. METODE PENELITIAN

2.1. Sumber Data

Data yang digunakan pada penelitian ini bersumber dari Kaggle, sebuah *platform* daring yang menyediakan dataset dari berbagai bidang ilmu. Dataset yang digunakan untuk pengembangan model identifikasi penyakit jantung menggunakan SVM terdiri dari 768 entri, 500 diantaranya adalah data negatif dan 268 diantaranya adalah data positif. Sehingga untuk mengatasi ketidakseimbangan data, maka dilakukan penggunaan *oversampling* menggunakan *Synthetic Minority Oversampling Technique* (SMOTE), dan dihasilkan data yang seimbang yaitu 500 data negatif dan juga 500 data positif. Data tersebut mencakup *pregnancies*, *glucose*, *blood pressure*, *skin thickness*, *insulin*, *body mass index*, *diabetes pedigree function*, *age*, dan *outcome* (sebagai output) yang selanjutnya disebut sebagai parameter dalam menentukan deteksi dini pada penyakit Diabetes Melitus

2.2. Tahapan penelitian



Gambar 1. Tahapan penerapan metode SVM

Pada gambar 1 diatas merupakan tahapan penelitian dalam pengembangan menggunakan metode *machine learning*, diantaranya sebagai berikut.

2.2.1. Pengumpulan Data

Data diambil dari data sekunder yang berasal dari situs kaggle, dimana data ini merupakan data mentah yang diolah dalam deteksi dini penyakit diabetes melitus. Data tersebut memiliki 9 parameter, dimana diantaranya 8 parameter sebagai input dan 1 parameter sebagai output diantaranya : *pregnancies*, *glucose*, *blood pressure*, *skin thickness*, *insulin*, *body mass index*, *diabetes pedigree function*, *age*, dan *outcome*

2.2.2. Pre-Processing Data

Data yang telah diambil selanjutnya diolah untuk menghasilkan data-data terbaik, beberapa tahapan pengolahannya termasuk pembersihan *outliers*, pengisian *missing data value*, penyesuaian data dan *labeling*, *normalisasi* data, pembagian *train* dan *test* data, serta penggunaan SMOTE. Berikut tahapan pada preprocessing data :

a. Pembersihan Outliers

Outliers adalah nilai yang sangat ekstrem atau tidak biasa dalam kumpulan data. Beberapa manfaat pembersihan outliers adalah mengurangi pengaruh outliers pada model, meningkatkan robustness, meningkatkan akurasi, dan stabilitas pada pelatihan. Adapun pembersihan outliers diimplementasikan pada setiap parameter yang digunakan dapat dilihat pada gambar 2 berikut ini.

Pembersihan Outliers

```
[9] data[data.BMI > (data.BMI.mean() + 3 * data.BMI.std())]
```

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
125	1.0	88.0	30.0	42.0	99.0	55.0	0.496	26	1.0
177	0.0	129.0	110.0	46.0	130.0	67.1	0.319	26	1.0
445	0.0	180.0	78.0	63.0	14.0	59.4	2.420	25	1.0
673	3.0	123.0	100.0	35.0	240.0	57.3	0.880	22	0.0

```
[10] data[data.BloodPressure > (data.BloodPressure.mean() + 3 * data.BloodPressure.std())]
```

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
43	9.0	171.0	110.0	24.0	240.0	45.4	0.721	54	1.0
106	1.0	96.0	122.0	NaN	NaN	22.4	0.207	27	0.0
177	0.0	129.0	110.0	46.0	130.0	67.1	0.319	26	1.0
549	4.0	189.0	110.0	31.0	NaN	28.5	0.680	37	0.0
691	13.0	158.0	114.0	NaN	NaN	42.3	0.257	44	1.0

```
[11] data[data.Insulin > (data.Insulin.mean() + 3 * data.Insulin.std())]
```

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
8	2.0	197.0	70.0	45.0	543.0	30.5	0.158	53	1.0
13	1.0	189.0	60.0	23.0	846.0	30.1	0.398	59	1.0
228	4.0	197.0	70.0	39.0	744.0	36.7	2.329	31	0.0
247	0.0	165.0	90.0	33.0	680.0	52.3	0.427	23	0.0
286	5.0	155.0	84.0	44.0	545.0	38.7	0.619	34	0.0
409	1.0	172.0	68.0	49.0	579.0	42.4	0.702	28	1.0
584	8.0	124.0	76.0	24.0	600.0	28.7	0.687	52	1.0
655	2.0	155.0	52.0	27.0	540.0	38.7	0.240	25	1.0

```
[12] data[data.SkinThickness > (data.SkinThickness.mean() + 3 * data.SkinThickness.std())]
```

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
445	0.0	180.0	78.0	63.0	14.0	59.4	2.420	25	1.0
579	2.0	197.0	70.0	99.0	NaN	34.7	0.575	62	1.0

Gambar 2. Pembersihan Outliers Pada Data

b. Pengisian *Missing Data Value*

Pengisian *missing data value* dilakukan dengan mengisi data yang hilang dengan mean dan median. Penentuan dalam mengisi *missing data value* ini sendiri diantaranya adalah mempertimbangkan beberapa hal, kolom yang diisi dengan nilai mean biasanya memiliki distribusi data yang cenderung normal dan memiliki data hilang yang tidak terlalu banyak. Sedangkan kolom yang diisi dengan nilai median biasanya memiliki distribusi data yang cenderung abnormal dan terkadang memiliki data hilang yang banyak. Terdapat beberapa manfaat pengisian *missing data value*, yaitu mempertahankan ukuran sampel, meningkatkan keterpakaian data, dan mengurangi bias. Gambar 3 berikut ini merupakan penerapan pengisian *missing data value*.

```
Pengisian Missing Data Value
```

```
data['Glucose'].replace(0, np.mean(data['Glucose']), inplace=True)
data['BloodPressure'].replace(0, np.mean(data['BloodPressure']), inplace=True)
data['SkinThickness'].replace(0, np.median(data['SkinThickness']), inplace=True)
data['Insulin'].replace(0, np.median(data['Insulin']), inplace=True)
data['BMI'].replace(0, np.mean(data['BMI']), inplace=True)
```

```
[9]
```

```
data.head()
```

```
[10]
```

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148.0	72.0	35	30.5	33.6	0.627	50	1
1	1	85.0	66.0	29	30.5	26.6	0.351	31	0
2	8	183.0	64.0	23	30.5	23.3	0.672	32	1
3	1	89.0	66.0	23	94.0	28.1	0.167	21	0
4	0	137.0	40.0	35	168.0	43.1	2.288	33	1

Gambar 3. Pengisian *Missing Data Value*

c. Normalisasi Data

Normalisasi data adalah suatu proses di mana nilai-nilai dalam dataset disesuaikan atau dinormalisasi agar memiliki skala yang seragam atau sebanding. Dimana data dari kolom yang diterapkan normalisasi adalah sebagai berikut *Pregnancies*, *Glucose*, *Blood Pressure*, *Skin Thickness*, *Insulin*, *BMI*, *Diabetes Pedigree Function*, *Age*. Normalisasi data memberikan beberapa keuntungan yaitu diantaranya meningkatkan konvergensi model, mengurangi pengaruh outliers, penyesuaian skala yang sama, meningkatkan kinerja model dan akurasi. Adapun penerapan normalisasi data diterapkan pada gambar 4 sebagai berikut.

```
Normalisasi Data
```

```
x = data.drop('Outcome', axis='columns')
y = data['Outcome']
scalar = ['Pregnancies', 'Glucose', 'BloodPressure', 'SkinThickness', 'Insulin', 'BMI', 'DiabetesPedigreeFunction', 'Age']

from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.preprocessing import MinMaxScaler

scale = MinMaxScaler()
x_scaled = scale.fit_transform(x[scalar])
```

Gambar 4. Penerapan Normalisasi Data

d. Penggunaan SMOTE

SMOTE merupakan salah satu metode dalam penyeimbangan data, dimana merupakan pengembangan dari metode oversampling. Terdapat beberapa manfaat penggunaan SMOTE diantaranya meningkatkan kinerja model, mencegah dominansi kelas mayoritas, meminimalkan risiko bias, dan penyimbangan model. Adapun penerapan penggunaan SMOTE pada gambar 5.

```
Penggunaan SMOTE
```

```
from imblearn.over_sampling import SMOTE

smote = SMOTE(sampling_strategy='minority')
x_sm, y_sm = smote.fit_resample(x_scaled, y)
```

```
[13]
```

Gambar 5. Penggunaan SMOTE

e. Pembagian *train dan test data*

Pembagian ini bertujuan untuk mengukur sejauh mana model yang dilatih pada set pelatihan dapat menggeneralisasi dengan baik pada data yang belum pernah dilihat sebelumnya, yang terdapat pada set pengujian. Adapun penerapan pada gambar 6 berikut ini.

Pembagian Train dan Test pada Data

```
x_train, x_test, y_train, y_test = train_test_split(x_scaled, y, test_size=0.2, random_state=30)

x_train_sm, x_test_sm, y_train_sm, y_test_sm = train_test_split(x_sm, y_sm, test_size=0.2, random_state=44)
```

Gambar 6. Pembagian Train dan Test

2.2.3. Analisis Model *Machine learning*

Data yang telah diolah selanjutnya dilakukan analisis tingkat lanjut, untuk menentukan algoritma yang tepat, ini digunakan dengan menerapkan penggunaan cross validation dan Grid Search validation. Adapun penerapan dapat dilihat pada gambar 7 berikut.

```
model_parameter = {
    'Linear_Regression': {
        'model': LinearRegression(),
        'parameter': {}
    },
    'Logistic_Regression': {
        'model': LogisticRegression(max_iter=100000),
        'parameter': {
            'C': [1, 10, 30, 100],
            'solver': ['lbfgs', 'liblinear', 'newton-cg', 'sag', 'saga']
        }
    },
    'Decision_Tree': {
        'model': DecisionTreeClassifier(),
        'parameter': {
            'criterion': ['gini', 'entropy']
        }
    },
    'Random_Forest': {
        'model': RandomForestClassifier(),
        'parameter': {
            'n_estimators': [50, 100, 150, 200]
        }
    },
    'SVM': {
        'model': SVC(gamma='scale'),
        'parameter': {
            'C': [1, 5, 10, 15, 20, 25, 30, 40, 50],
            'kernel': ['rbf', 'poly']
        }
    },
    'GaussianNB': {
        'model': GaussianNB(),
        'parameter': {}
    },
    'MultinomialNB': {
        'model': MultinomialNB(),
        'parameter': {}
    },
    'Kneighbors': {
        'model': KNeighborsClassifier(),
        'parameter': {
            'n_neighbors': [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
        }
    }
}
```

Gambar 7. Parameter Data

2.2.4. Implementasi *Machine learning*

Data yang telah dianalisis tingkat lanjut selanjutnya diolah menggunakan algoritma terbaik untuk data, dalam konteks ini adalah SVM diterapkan beserta parameter terbaiknya. Berikut gambar 8 merupakan implementasinya.

```
using_model = SVC(C=15, kernel='poly')
model_2 = BaggingClassifier(
    using_model,
    n_estimators=500,
    max_samples=0.8,
    oob_score=True
)
model_2.fit(x_train_sm, y_train_sm)

joblib.dump(model_2, 'D:/Source Code/Python/model/diabetes_model.pkl')
```

Gambar 8. Implementasi Model *Machine learning* Algoritma SVM

2.2.5. Pemanfaatan Model

Model *machine learning* yang telah dilatih dan disimpan, selanjutnya dapat digunakan untuk pengolahan data lebih lanjut, dalam konteks ini dirancang visualisasi data dengan mengembangkan *website* sederhana untuk memanfaatkan model..

3. HASIL DAN PEMBAHASAN

3.1. Hasil implementasi model SVM

Melalui preprocessing data didapatkan data model yang seimbang dengan perbandingan 500 data negative dan 500 data positif melalui menggunakan oversampling. Serta dilakukan pengecekan pada setiap parameter. Dari 9 parameter terdapat beberapa parameter yang kosong diantaranya *skin thickness*, *insulin*, *body mass index* maka dilakukan penggantian data hilang dengan mean yang dapat mempertahankan jumlah sampel data serta menghindari kehilangan informasi yang saling berhubungan antara satu parameter dengan parameter yang lainnya. Berikut adalah *classification report* dari SVM. Pada model yang dapat dilihat keseluruhannya, diukur *precision*, *recall*, *f1-score*, *accuracy*, dan lainnya merupakan ukuran kinerja untuk model Support Vector Machine (SVM) yang digunakan dalam penelitian ini. Hasil pengukuran kinerja dapat dilihat dari gambar 9 dibawah ini.

	precision	recall	f1-score	support
0	0.82	0.80	0.81	99
1	0.81	0.83	0.82	101
accuracy			0.81	200
macro avg	0.82	0.81	0.81	200
weighted avg	0.82	0.81	0.81	200

Gambar 9. Classification Report SVM

Keterangan:

a. Akurasi (*Accuracy*)

Akurasi mengukur sejauh mana model dapat memprediksi dengan benar, dihitung sebagai rasio prediksi yang benar (*True Positive* dan *True Negative*) terhadap total pengamatan. Akurasi yang dihasilkan pada SVM adalah 81 persen.

b. Presisi (*Precision*)

Presisi mengukur sejauh mana prediksi positif model adalah benar, dihitung sebagai rasio True Positive terhadap total prediksi positif (*True Positive* dan *False Positive*). Presisi pada hasil 0 (*Negative Diabetes melitus*) adalah 82 persen, 1 (*Positive Diabetes melitus*) adalah 81 persen.

c. Recall (*Sensitivitas atau True Positive Rate*)

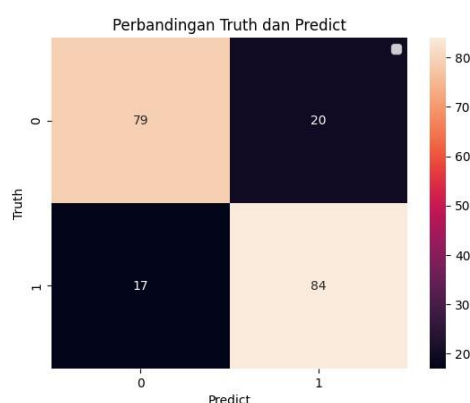
Recall mengukur sejauh mana model dapat mengidentifikasi semua kasus positif, dihitung sebagai rasio True Positive terhadap total kasus positif (True Positive dan False Negative). Recall pada hasil 0 (Negative Diabetes melitus) adalah 80 persen, 1 (Positive Diabetes melitus) adalah 83 persen

d. F1-Score

F1-Score merupakan perpaduan antara presisi dan recall, memberikan gambaran holistik tentang kinerja model. F-1 Score pada hasil 0 (Negative Diabetes melitus) adalah 81 persen, 1 (Positive Diabetes melitus) adalah 82 persen

e. Confusion Matrix

Matriks kebingungan memberikan gambaran detail tentang kinerja model, dengan menyajikan jumlah True Positive, True Negative, False Positive, dan False Negative yang dapat dilihat pada gambar 10 berikut ini.



Gambar 10. Confusion Matrix Model

Hasil evaluasi mencakup nilai-nilai metrik di atas, serta interpretasi dari metrik tersebut dalam konteks identifikasi penyakit diabetes. Setelah dilakukan pengujian, didapatkan bahwa model SVM lebih cenderung optimal karena dapat menangani dimensi data yang kompleks dan juga menghasilkan representasi hasil output yang cenderung seimbang. *Hyper parameter* yang digunakan juga telah disesuaikan dicocokkan, sehingga didapatkan model yang paling optimal dan baik dalam menangani kasus identifikasi penyakit diabetes melitus. Metrik-metrik ini membantu mengevaluasi kinerja model SVM dalam mengklasifikasikan apakah seseorang mungkin menderita diabetes (nilai 1) atau tidak (nilai 0). Nilai-nilai ini menunjukkan bahwa model memiliki kinerja yang cukup baik dengan akurasi keseluruhan sebesar 81%, dan precision, recall, serta F1-scores menunjukkan kinerja yang seimbang pada setiap nilai (nilai Positif dan nilai negatif). Hal ini menunjukkan bahwa model SVM dapat diandalkan sebagai alat untuk deteksi dini diabetes, yang sangat penting untuk pencegahan dan pengelolaan penyakit ini.

3.2. Perbandingan Algoritma

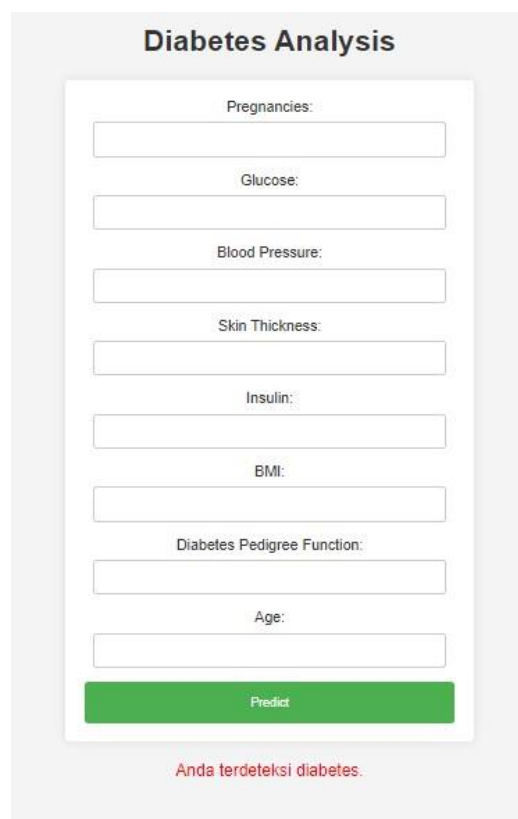
Implementasi pengelolaan data untuk deteksi dini penyakit diabetes melitus juga dilakukan pada beberapa penggunaan algoritma lain menggunakan metode GridSearchCV yang merupakan pengembangan dari metode *K-Fold Validation*. Perbandingan ini mengukur dengan melatih model menggunakan algoritma dan parameter yang berbeda selama 5 kali (tiap model dan parameter), kemudian diukur nilai mean dari data tersebut. Algoritma yang digunakan diantaranya : *linear regression*, *logistic Regression*, *decision tree*, *Random Forest*, *SVM*, *Gaussian*, *Multinomial*, *k-neighbors*. Skor terbaik menunjukkan seberapa baik setiap model dalam melakukan prediksi berdasarkan data yang digunakan. Model dengan skor tertinggi adalah Logistic Regression dan Random Forest, dan SVM dengan skor diatas 0.76 yang ditampilkan pada gambar 11 berikut ini.

	Nama Model	Best Parameter	Best Scores
0	Linear Regression	{}	0.296800
1	Logistic Regression	{'C': 10, 'solver': 'lbfgs'}	0.769570
2	Decision Tree	{'criterion': 'gini'}	0.711035
3	Random Forest	{'n_estimators': 50}	0.769570
4	SVM	{'C': 15, 'kernel': 'poly'}	0.768245
5	GaussianNB	{}	0.750047
6	MultinomialNB	{}	0.648400
7	Kneighbors	{'n_neighbors': 10}	0.752661

Gambar 11. Perbandingan Masing-Masing Model dengan GridSearchCV n.

3.3. Output

Website statis sederhana digunakan dalam mevisualisasikan analisis penggunaan algoritma dalam deteksi dini penyakit diabetes melitus. Model *machine learning* dan *website* statis ini nantinya dapat digunakan secara terbuka untuk kepentingan medis . *Website* statis sederhana yang telah dibuat dapat dijalankan pada komputer pribadi dan juga telah diterapkan model *machine learning* sederhana didalamnya dengan menginputkan 8 parameter yang dapat dilihat pada gambar 12 berikut ini.



Gambar 12. *Website* Deteksi Diabetes

4. KESIMPULAN

Penelitian ini telah berjalan sesuai dengan diagram alir yang direncanakan dimulai dari pengambilan dataset, *preprocessing* data, analisis data, implementasi *machine learning* dan interpretasi hasil. Dari hasil pengolahan 8 parameter diantaranya *pregnancies*, *glucose*, *blood pressure*, *skin thickness*, *insulin*, *body mass index*, *diabetes pedigree function*, *age* dengan 768 entri menggunakan model machine learning SVM didapatkan tingkat akurasi sebanyak 81%. Selain itu model SVM lebih

cenderung optimal karena dapat menangani dimensi data yang kompleks dan juga menghasilkan representasi hasil output yang cenderung seimbang serta model SVM didapat sebagai model yang paling optimal dan baik dalam menangani kasus identifikasi penyakit diabetes melitus.

DAFTAR PUSTAKA

- Biologi, J., Sains dan Teknologi, F., Alauddin Makassar, U., Pemeriksaan, C., Pengobatan dan Cara Pencegahan LESTARI, C., Aisyah Sijid, S., Studi Biologi, P., & Alauddin Makassar Jl Yasin Limpo Gowa, U. H. (n.d.). *Diabetes Melitus: Review Etiologi*. <http://journal.uin-alauddin.ac.id/index.php/psb>
- Fadli, S., Ashari, M., Studi Sistem Informasi, P., & Lombok, S. (2021). *JISA (Jurnal Informatika dan Sains) Optimization of Support Vector Machine Method Using Feature Selection to Improve Classification Results*.
- Gede, A., Pradnyana, S., Kom, M., Kom, K., Agustini, S., & Si, M. S. (n.d.). *Konsep Dasar Data Mining*.
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining. Concepts and Techniques, 3rd Edition (The Morgan Kaufmann Series in Data Management Systems)*.
- Huda Ovirianti, N., Zarlis, M., & Mawengkang, H. (2022). Support Vector Machine Using A Classification Algorithm. *Jurnal Dan Penelitian Teknik Informatika*, 6(3). <https://doi.org/10.33395/sinkron.v7i3>
- Jun, Z. (2021). The Development and Application of Support Vector Machine. *Journal of Physics: Conference Series*, 1748(5). <https://doi.org/10.1088/1742-6596/1748/5/052006>
- Kurniawaty, E. (2014). Diabetes Mellitus. In *Evi Kurniawaty JUKE* (Vol. 4, Issue 7).
- Nurmay Stiani, S., Nur Sabilla, A., Korespondensi, P., oleh, D., Tinggi Ilmu Kesehatan Salsabila, S., & Banten, S. (2022). Pengaruh Pelayanan Kefarmasian di Rumah terhadap Pengendalian Kadar Glukosa Darah Pasien Diabetes Mellitus Tipe 2 di Puskesmas Ngampilan Yogyakarta. *Jurnal Ilmiah Kesehatan Delima*, 4(2), 74–77. <https://doi.org/10.60010/jikd/v4i2.79>
- Tan, P.-N., Steinbach, M., & Kumar, V. (n.d.). *Introduction to data mining*.
- Tomic, D., Shaw, J. E., & Magliano, D. J. (2022). The burden and risks of emerging complications of diabetes mellitus. In *Nature Reviews Endocrinology* (Vol. 18, Issue 9, pp. 525–539). Nature Research. <https://doi.org/10.1038/s41574-022-00690-7>
- Veronica Agustin, A., Voutama, A., Singaperbangsa Karawang HS Ronggo Waluyo, U. J., & Barat, J. (2023). Implementasi Data Mining Klasifikasi Penyakit Diabetes Pada Perempuan Menggunakan Naïve Bayes. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 7, Issue 2)..