

## Hybrid XGBoost-SVM Model untuk Sistem Pendukung Keputusan dalam Prediksi Penyakit Diabetes

Muhammad Surono<sup>1</sup>, Muhammad Fadli<sup>2</sup>, Dian Sri Purwamti<sup>3</sup>, Erliyan Redy Susanto<sup>4</sup>

<sup>1,3,4</sup>Magister Ilmu Komputer, Universitas Teknokrat Indonesia, Kota Bandar Lampung, Indonesia

<sup>2</sup>Jurusan Ekonomi dan Bisnis, Politeknik Negeri Lampung, Kota Bandar Lampung, Indonesia

Email: <sup>1</sup>muhamdsurono@teknokrat.ac.id, <sup>2</sup>muhammadfadiofficial@polinela.ac.id,

<sup>3</sup>dian\_sri\_purwanti@teknokrat.ac.id, <sup>4</sup>erliyan.redy@teknokrat.ac.id

### Abstract

*Diabetes is a chronic disease that continues to rise globally each year, requiring early detection for more effective prevention. This study develops an artificial intelligence-based decision support system for diabetes prediction using a Hybrid XGBoost-SVM model. The model combines the Support Vector Machine (SVM), known for its interpretability, with XGBoost (XGB), which enhances accuracy through boosting techniques. The study utilizes the Pima Indians Diabetes Dataset, undergoing preprocessing, normalization, data splitting, and model training. The evaluation compares accuracy, precision, recall, and F1-score across the three models. Experimental results indicate that XGBoost and SVM both achieve an accuracy of 75%. However, the Hybrid XGBoost-SVM model provides consistently improved performance, achieving the highest accuracy (77%), along with increased precision (70%) and F1-score (65%). Although the numerical improvement in accuracy appears relatively small, this enhancement is significant in the medical context, especially due to improved precision and balanced classification. This study concludes that the Hybrid XGBoost-SVM approach offers a more optimal and reliable alternative in decision support systems for diabetes prediction. Future research can explore other model combinations, such as Stacking or Weighted Voting, to enhance predictive performance further.*

**Keywords:** Diabetes Prediction, Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), Hybrid XGBoost-SVM Model, Decision Support System (DSS).

### Abstrak

Diabetes merupakan salah satu penyakit kronis yang terus meningkat setiap tahunnya dan memerlukan deteksi dini untuk pencegahan yang lebih efektif. Dalam penelitian ini, dikembangkan sistem pendukung keputusan berbasis kecerdasan buatan untuk prediksi penyakit diabetes menggunakan model *Hybrid XGBoost-SVM*. Model ini menggabungkan *Support Vector Machine (SVM)* yang memiliki interpretabilitas tinggi dengan *Extreme Gradient Boosting (XGBoost)* yang mampu meningkatkan akurasi melalui teknik boosting. Penelitian dilakukan menggunakan Pima Indians Diabetes Dataset, melalui tahapan pra-pemrosesan data, normalisasi, pembagian data latih dan uji, serta pelatihan model. Evaluasi dilakukan dengan membandingkan akurasi, presisi, recall, dan F1-score dari masing-masing model. Hasil eksperimen menunjukkan bahwa model *Hybrid XGBoost-SVM* memberikan performa terbaik, dengan peningkatan akurasi dari 75% (*XGBoost/SVM*) menjadi 77%, serta peningkatan presisi dan F1-score dibandingkan dua model tunggal. Meskipun peningkatannya terlihat moderat secara angka, dalam konteks deteksi penyakit, terutama untuk mengurangi false negative, peningkatan ini dianggap signifikan secara praktis karena dapat meningkatkan ketepatan diagnosis pasien. Penelitian ini menyimpulkan bahwa pendekatan *Hybrid XGBoost-SVM* dapat menjadi solusi alternatif yang efektif dalam sistem pendukung keputusan medis, terutama untuk deteksi dini diabetes. Penelitian selanjutnya dapat mengeksplorasi kombinasi model lain seperti Stacking atau Weighted Voting untuk peningkatan lebih lanjut.

**Kata Kunci:** Prediksi Diabetes, *Support Vector Machine (SVM)*, *Extreme Gradient Boosting (XGBoost)*, *Hybrid XGBoost-SVM Model*, Sistem Pendukung Keputusan (SPK).

## 1. PENDAHULUAN

Diabetes merupakan penyakit kronis yang terus mengalami peningkatan jumlah kasus secara global setiap tahunnya, sehingga prediksi dini menjadi aspek krusial dalam upaya pencegahan dan pengendalian (Oktavia, Aeni and Saraswati, 2020). Salah satu faktor risiko yang perlu diperhatikan oleh penderita diabetes adalah konsumsi makanan dengan indeks glikemik tinggi, yang dapat memicu lonjakan insulin berlebihan dan mengganggu kestabilan kadar gula darah. Ketidakseimbangan ini, apabila terjadi secara berulang, berpotensi menyebabkan stres metabolismik dan kerusakan organ secara bertahap. Di sisi lain, diabetes juga digolongkan sebagai penyakit metabolismik yang ditandai oleh peningkatan kadar glukosa dalam darah akibat gangguan sekresi atau resistensi insulin (Erlin *et al.*, 2022). Oleh karena itu, teknologi yang mampu mendeteksi potensi diabetes secara dini sangat dibutuhkan dalam praktik klinis.

Salah satu solusi yang dapat diterapkan adalah Sistem Pendukung Keputusan (SPK) atau *Computer-Based Decision Support System (DSS)*. Teknologi ini dirancang untuk membantu pengambilan keputusan kompleks secara lebih efektif dengan mempertimbangkan berbagai alternatif dan kriteria (Debora Mait *et al.*, 2022) (Wahyuni, 2025). Dalam dunia medis, SPK yang berbasis kecerdasan buatan telah banyak dikembangkan untuk menganalisis pola data pasien, mengidentifikasi faktor risiko, dan menghasilkan rekomendasi diagnostik yang akurat dan efisien (Lestari, Nadhiroh and Novia, 2021) (Karyadiputra and Setiawan, 2023). Selain itu, solusi berbasis Sistem Pendukung Keputusan (SPK) memiliki kemampuan untuk menangani permasalahan dengan kondisi semi terstruktur maupun tidak terstruktur melalui mekanisme pemecahan masalah dan komunikasi yang efektif (Agustanti, Alawiyah and Suhandi, 2021).

Algoritma *Support Vector Machine (SVM)* merupakan salah satu teknik pembelajaran mesin yang efektif dalam tugas klasifikasi, khususnya untuk data berdimensi tinggi. *Support Vector Machine (SVM)* bekerja dengan mencari *hyperplane* optimal sebagai pemisah antar kelas dalam ruang fitur, dan efektif menangani data *non-linear* melalui penggunaan kernel (Amelia, Indra and Masruriyah, 2022) (Wibowo, Indarti and Laraswati, 2024). *Dataset* dibersihkan dan dinormalisasi terlebih dahulu sehingga siap untuk dimasukkan ke dalam model *Support Vector Machine (SVM)*. Model *Support Vector Machine (SVM)* diproses dan diuji sehingga mendapatkan model terbaik untuk melakukan diagnosis. Keluaran dari model *Support Vector Machine (SVM)* akan mendiagnosis pasien yang menderita diabetes ataupun yang tidak menderita diabetes (Mucholladin, Bachtiar and Furqon, 2021). Sementara itu, *Extreme Gradient Boosting (XGBoost)* adalah algoritma *ensemble* berbasis *decision tree* yang dibangun secara bertahap melalui *boosting*, dan dikenal karena efisiensinya dalam meningkatkan akurasi serta kemampuannya dalam menangani *overfitting* dan pemilihan fitur secara otomatis (Rayadin *et al.*, 2024) (Syukron, Santoso and Widiharih, 2020) (Ramadhanti *et al.*, 2024). Pohon keputusan akan diperbaiki setiap penambahan pohon dengan mempertimbangkan kesalahan prediksi yang dilakukan oleh model sebelumnya. *Extreme Gradient Boosting (XGBoost)* memiliki performa yang baik dalam klasifikasi, bahkan mengungguli beberapa algoritma lain seperti *Naïve Bayes* dan *Random Forest* dalam konteks tertentu (Soelistijadi *et al.*, 2024) (Khairunnisa, 2023). *Extreme Gradient Boosting (XGBoost)* juga merupakan algoritma berbasis pohon yang lebih efisien dan Scalable (Nasution, Saedudin and Widartha, 2021). Algoritma *Extreme Gradient Boosting (XGBoost)* dapat mengatasi data medis yang kompleks dan beragam, dan dapat memenuhi persyaratan ketepatan waktu dan akurasi diagnosis tambahan dengan lebih baik (Andryan, Fajri and Sulistyowati, 2022).

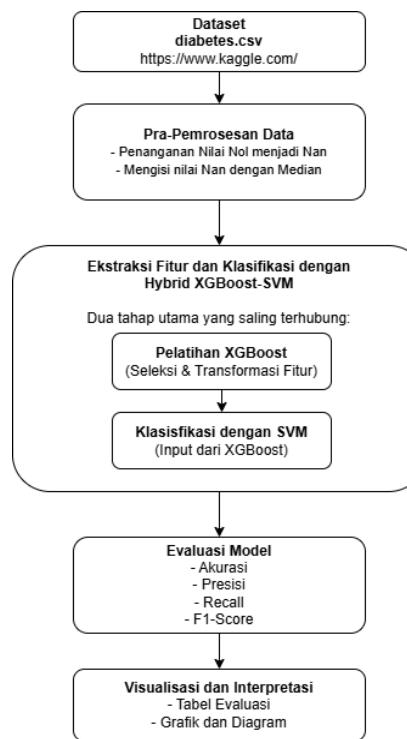
Dalam penelitian ini dikembangkan model *Hybrid XGBoost-SVM* dengan

menggabungkan keunggulan kedua algoritma tersebut untuk meningkatkan performa klasifikasi dalam sistem pendukung keputusan prediksi diabetes. Model ini tidak hanya menghasilkan prediksi berdasarkan *machine learning*, tetapi juga dapat memperkuat proses pengambilan keputusan medis secara cepat dan akurat. Berdasarkan hasil eksperimen terhadap Pima Indians Diabetes Dataset, model *hybrid* menunjukkan kinerja yang lebih stabil dibandingkan model tunggal, baik dari sisi akurasi maupun keseimbangan klasifikasi (Abdurrosyid, Teguh and Almias, 2025) (Kurnia *et al.*, 2023). Namun, untuk mengoptimalkan keakuratan model lebih lanjut, pendekatan *hybrid* dengan kombinasi *Voting Classifier* diterapkan guna memperoleh prediksi yang lebih stabil dan akurat.

Penelitian ini bertujuan untuk membangun dan mengevaluasi model *Hybrid XGBoost-SVM* dalam sistem pendukung keputusan untuk prediksi penyakit diabetes, serta membandingkan performanya dengan masing-masing metode tunggal. Model ini diharapkan dapat meningkatkan akurasi klasifikasi dan keandalan prediksi dalam mendukung deteksi dini risiko diabetes oleh tenaga medis. Studi sebelumnya juga menunjukkan efektivitas model ini pada kasus medis lainnya seperti prediksi penyakit jantung hipertensi (Chang *et al.*, 2019).

## 2. METODOLOGI PENELITIAN

Untuk Proses komputasi menggunakan bahasa python dengan menggunakan *google colab*. Adapun *library python* yang digunakan antara lain: *pandas*, *numpy*, *scikit-learn*, dan *matplotlib*. Dataset yang disiapkan berisikan data *Pregnancies*, *Glucose*, *BloodPressure*, *SkinThickness*, *Insulin*, *BMI* (*Indeks massa tubun*), *DiabetesPedigreeFunction*, *Age*, *Outcome* (hasil). Dataset yang digunakan sebanyak 768 baris dan 9 kolom. Penelitian ini menggunakan pendekatan kuantitatif dengan metode pengolahan data secara sistematis untuk membangun model prediksi penyakit diabetes menggunakan algoritma *Hybrid XGBoost-SVM*. Adapun tahapan-tahapan dalam metode penelitian ini adalah sebagai berikut:



Gambar 1. Diagram alir penelitian

## 2.1 Dataset

*Dataset* yang digunakan adalah Pima Indians Diabetes *Dataset* yang terdiri dari 768 baris dan 9 kolom, dengan 8 fitur input dan 1 label output (*Outcome*). Label bernilai 1 menunjukkan pasien menderita diabetes, sedangkan nilai 0 menunjukkan sebaliknya.

Variabel dataset dapat dilihat pada gambar dibawah ini:

```
→ Dataset memiliki 768 baris dan 9 kolom.
Missing values:
Pregnancies          0
Glucose              0
BloodPressure        0
SkinThickness        0
Insulin              0
BMI                  0
DiabetesPedigreeFunction 0
Age                  0
Outcome              0
dtype: int64
/usr/local/lib/python3.11/dist-packages/xgboost/c
Parameters: { "use_label_encoder" } are not used.
```

Gambar 2. Variabel *dataset*

Lima baris data teratas dataset dapat dilihat pada gambar dibawah ini:

```
5 baris data teratas:
   Pregnancies  Glucose  BloodPressure  SkinThickness  Insulin  BMI  \
0           6      148            72            35       0  33.6
1           1       85            66            29       0  26.6
2           8      183            64            0       0  23.3
3           1       89            66            23      94  28.1
4           0      137            40            35     168  43.1

   DiabetesPedigreeFunction  Age  Outcome
0           0.627    50       1
1           0.351    31       0
2           0.672    32       1
3           0.167    21       0
4           2.288    33       1
```

Gambar 3. Lima data teratas

Lima baris data terbawah dataset dapat dilihat pada gambar dibawah ini:

```
5 baris data terbawah:
   Pregnancies  Glucose  BloodPressure  SkinThickness  Insulin  BMI  \
763          10      101            76            48      180  32.9
764          2       122            70            27       0  36.8
765          5      121            72            23      112  26.2
766          1      126            60            0       0  30.1
767          1       93            70            31       0  30.4

   DiabetesPedigreeFunction  Age  Outcome
763           0.171    63       0
764           0.340    27       0
765           0.245    30       0
766           0.349    47       1
767           0.315    23       0
```

Gambar 4. Lima data terbawah

## 2.2 Pra-pemrosesan Data

Tahapan ini bertujuan untuk membersihkan dan mempersiapkan data agar dapat digunakan dalam pelatihan model:

1. Penanganan nilai nol pada kolom medis seperti *Glucose*, *BloodPressure*, *SkinThickness*, *Insulin*, dan *BMI* yang tidak masuk akal jika bernilai nol. Nilai-nilai ini akan digantikan dengan nilai median kolom terkait.
2. Normalisasi data dengan metode *Min-Max Scaling* untuk menyetarakan skala fitur.
3. Pembagian data menjadi data latih (80%) dan data uji (20%).

## 2.3 Ekstraksi Fitur dan Klasifikasi dengan *Hybrid XGBoost-SVM*

Model *Hybrid XGBoost-SVM* dibangun dengan dua tahap utama yang saling terhubung:

- 1) Extreme Gradient Boosting (XGBoost) sebagai Ekstraktor Fitur dan Selektor Atribut Penting
  1. Model *Extreme Gradient Boosting (XGBoost)* dilatih terlebih dahulu pada data latih.
  2. *Extreme Gradient Boosting (XGBoost)* digunakan untuk menghitung *feature importance* berdasarkan nilai *gain*.
  3. Berdasarkan hasil seleksi, fitur-fitur penting digunakan untuk mentransformasi data latih dan uji menjadi vektor fitur baru, yang selanjutnya digunakan sebagai input ke model *Support Vector Machine (SVM)*.
- 2) Support Vector Machine (SVM) sebagai Model Klasifikasi
  1. Vektor fitur hasil transformasi dari *Extreme Gradient Boosting (XGBoost)* menjadi input untuk algoritma *Support Vector Machine (SVM)*.
  2. *Support Vector Machine (SVM)* digunakan sebagai tahap klasifikasi akhir, bukan sebagai metode ekstraksi fitur terpisah.
  3. Konfigurasi parameter *Support Vector Machine (SVM)*:
    - a. *Kernel: Radial Basis Function (RBF)*
    - b. *C (Regularization): 1.0*
    - c. *Gamma: 'scale' (default sklearn)*
    - d. *Random State: 42*
  4. Fungsi *SVC* dari pustaka *Scikit-learn* digunakan dalam proses pelatihan dan pengujian model.

## 2.4 Evaluasi Model

Tahap terakhir adalah mengevaluasi performa dari ketiga model:

1. *Extreme Gradient Boosting (XGBoost)* (sebagai *baseline* pertama)
2. *Support Vector Machine (SVM)* (sebagai *baseline* kedua)
3. *Hybrid XGBoost-SVM* (model utama)

Model dievaluasi menggunakan metrik klasifikasi:

1. *Akurasi* : proporsi prediksi benar terhadap seluruh data.
2. *Presisi* : proporsi prediksi positif yang benar-benar positif.
3. *Recall* : proporsi kasus positif yang berhasil dikenali.
4. *F1-Score* : harmonisasi antara *presisi* dan *recall*.

Hasil menunjukkan bahwa model *Hybrid XGBoost-SVM* memberikan akurasi tertinggi, yaitu 77%, dibandingkan dengan *Extreme Gradient Boosting (XGBoost)* (75%) dan *Support Vector Machine (SVM)* (75%). *Presisi* model *Hybrid XGBoost-SVM* juga

lebih tinggi (70%) dibandingkan dua model lainnya. Nilai *recall* model *Hybrid XGBoost-SVM* tetap sama dengan *Support Vector Machine (SVM)* (61%), namun *F1-score* meningkat menjadi 65%, menandakan model lebih seimbang dalam mengenali kelas positif dan negatif.

## 2.5 Visualisasi dan Interpretasi

Disajikan dalam bentuk tabel evaluasi dan grafik, Visualisasi fitur penting dari *Extreme Gradient Boosting (XGBoost)* untuk interpretasi medis lebih lanjut.

# 3. HASIL DAN PEMBAHASAN

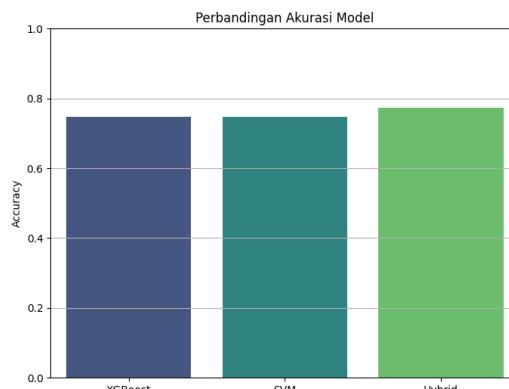
## 3.1 Hasil Penelitian

Distribusi jumlah pasien berdasarkan status diabetes pada Pima Indians Diabetes Dataset menunjukkan adanya ketidakseimbangan kelas yang cukup jelas. Dari grafik yang dihasilkan, terlihat bahwa jumlah pasien yang tidak menderita diabetes (label 0) secara signifikan lebih tinggi dibandingkan dengan pasien yang menderita diabetes (label 1). Tercatat terdapat sekitar 500 pasien *non-diabetes*, sedangkan pasien yang teridentifikasi positif diabetes hanya berjumlah sekitar 268 orang. Ini berarti bahwa sekitar 69,5% dari total populasi tidak menderita diabetes, sementara sisanya 30,5% merupakan penderita diabetes.

Ketimpangan proporsi ini merupakan indikasi adanya *class imbalance* dalam *dataset*, yang menjadi perhatian penting dalam proses pelatihan model klasifikasi. Jika tidak ditangani dengan baik, model cenderung akan lebih akurat dalam mengenali kelas mayoritas (*non-diabetes*), tetapi kurang mampu mengenali kelas minoritas (*diabetes*). Hal ini dapat menyebabkan *false negative* rate yang tinggi, di mana pasien yang sebenarnya positif diabetes justru diprediksi negatif, yang tentu berisiko besar terhadap aspek kesehatan masyarakat.

Meskipun dataset yang digunakan memiliki ketidakseimbangan kelas, dalam penelitian ini tidak diterapkan teknik penyeimbangan data seperti *SMOTE*, *oversampling*, atau penyesuaian bobot kelas. Hal ini dilakukan untuk mempertahankan distribusi data asli sebagaimana kondisi nyata di lapangan medis. Namun, konsekuensinya adalah adanya potensi bias klasifikasi terhadap kelas mayoritas, yang tercermin dari nilai *recall* yang belum maksimal untuk kelas positif. Implikasi dari ketidakseimbangan ini juga diperhatikan dalam interpretasi hasil dan disarankan untuk ditangani pada penelitian selanjutnya.

## 3.2 Perbandingan akurasi model *Extreme Gradient Boosting (XGBoost)*, *Support Vector Machine (SVM)* dan *Hybrid XGBoost-SVM*

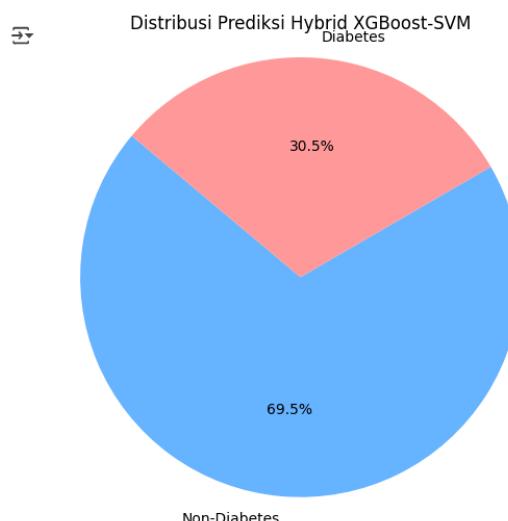


Gambar 5. Bar chart dari perbandingan akurasi model

Bar chart ini menyajikan perbandingan akurasi dari tiga model yang berbeda, yaitu *Extreme Gradient Boosting (XGBoost)*, *SVM*, dan model *Hybrid XGBoost-SVM*. Akurasi diukur dalam rentang 0 hingga 1, di mana 1 menunjukkan akurasi sempurna (semua prediksi benar) dan 0 menunjukkan akurasi terburuk (semua prediksi salah).

1. Hasil evaluasi menunjukkan bahwa model *Extreme Gradient Boosting (XGBoost)* dan *Support Vector Machine (SVM)* memiliki akurasi yang sama, yaitu 0.75 (75%), berdasarkan data uji yang digunakan.
2. *Hybrid XGBoost-SVM*: Model *Hybrid* yang menggabungkan *Extreme Gradient Boosting (XGBoost)* dan *Support Vector Machine (SVM)*, mencapai akurasi tertinggi di antara ketiganya yaitu 0,77 (77%). Hal ini menunjukkan bahwa penggabungan kedua metode ini meningkatkan kinerja prediksi secara keseluruhan.

### 3.3 Distribusi Prediksi *Hybrid XGBoost-SVM*



Gambar 6. Pie chart dari Distribusi Prediksi Hybrid XGBoost-SVM

Pie chart ini menggambarkan distribusi hasil prediksi dari model *Hybrid XGBoost-SVM* terhadap dua kelas: "Non-Diabetes" dan "Diabetes".

1. *Non-Diabetes*: Sebesar 69.5% dari total prediksi oleh model adalah pasien yang diprediksi tidak menderita diabetes.
2. *Diabetes*: Sebesar 30.5% dari total prediksi oleh model adalah pasien yang diprediksi menderita diabetes.

### 3.4 Hasil Perbandingan Model *Extreme Gradient Boosting (XGBoost)*, *Support Vector Machine (SVM)*, dan *Hybrid XGBoost-SVM*

Tabel 1. Hasil Perbandingan

Model	Akurasi	Presisi	Recall	F1-Score
<i>Extreme Gradient Boosting (XGBoost)</i>	0.75	0.65	0.59	0.62
<i>Support Vector Machine (SVM)</i>	0.75	0.65	0.61	0.63
<i>Hybrid XGBoost-SVM</i>	0.77	0.70	0.61	0.65

Dalam penelitian ini, tiga metode klasifikasi diterapkan untuk memprediksi *outcome* pasien diabetes, yaitu *Extreme Gradient Boosting (XGBoost)*, *Support Vector Machine (SVM)*, dan *Hybrid XGBoost-SVM*. Evaluasi dilakukan menggunakan empat metrik utama, yaitu *akurasi*, *presisi*, *recall*, dan *F1-score*.

### 3.4.1 Model *Extreme Gradient Boosting (XGBoost)*

Model *Extreme Gradient Boosting (XGBoost)* menunjukkan akurasi sebesar 0.75, dengan presisi 0.65, recall 0.59, dan *F1-score* 0.62. Meskipun model ini cukup baik dalam melakukan klasifikasi secara umum, nilai *recall* yang lebih rendah mengindikasikan bahwa masih ada cukup banyak kasus positif (diabetes) yang tidak berhasil dikenali oleh model.

### 3.4.2 Model *Support Vector Machine (SVM)*

Model *Support Vector Machine (SVM)* juga menghasilkan akurasi sebesar 0.75, namun menunjukkan sedikit perbaikan pada *recall* (0.61) dan *F1-score* (0.63). Ini menunjukkan bahwa *Support Vector Machine (SVM)* lebih baik dalam menangkap kasus positif dibandingkan *Extreme Gradient Boosting (XGBoost)*, meskipun presisi tetap sama (0.65). Dengan kata lain, *Support Vector Machine (SVM)* lebih sensitif terhadap kelas positif dibandingkan *Extreme Gradient Boosting (XGBoost)*.

### 3.4.3 Model *Hybrid XGBoost-SVM*

Model *ensemble Hybrid XGBoost-SVM* memberikan hasil terbaik di antara ketiga metode, dengan akurasi mencapai 0.77, presisi 0.70, recall 0.61, dan *F1-score* 0.65. Peningkatan presisi dan *F1-score* menunjukkan bahwa penggabungan dua model ini mampu mengurangi kesalahan prediksi positif dan meningkatkan keseimbangan antara presisi dan *recall*.

## 3.5 Evaluasi Kinerja Model *Extreme Gradient Boosting (XGBoost)*, *Support Vector Machine (SVM)*, dan *Hybrid XGBoost-SVM*

Tabel 2. Hasil Evaluasi Kinerja

Model	Akurasi	Presisi	Recall	<i>F1-Score</i>
<i>Extreme Gradient Boosting (XGBoost)</i>	75%	65%	59%	62%
<i>Support Vector Machine (SVM)</i>	75%	65%	61%	63%
<i>Hybrid XGBoost-SVM</i>	77%	70%	61%	65%

Model *Extreme Gradient Boosting (XGBoost)* menghasilkan akurasi sebesar 75% dengan *presisi* 65% dan *recall* 59%. Nilai *F1-score* sebesar 62% menunjukkan bahwa model memiliki keseimbangan sedang antara presisi dan *recall*, namun performanya dalam mengenali kelas positif (*recall*) masih tergolong rendah. Model *Support Vector Machine (SVM)* menunjukkan kinerja yang hampir setara dengan *Extreme Gradient Boosting (XGBoost)* dari sisi akurasi (75%) dan presisi (65%), namun sedikit lebih baik dalam mengenali kelas positif dengan *recall* sebesar 61%. *F1-score* juga mengalami peningkatan menjadi 63%, yang menandakan perbaikan dalam keseimbangan klasifikasi.

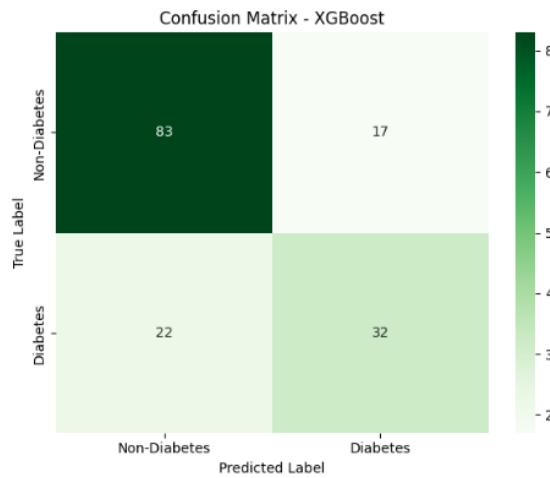
Dari kedua model menunjukkan bahwa terdapat perbedaan kecil pada aspek lain, seperti; *Support Vector Machine (SVM)* menunjukkan nilai *recall* yang lebih tinggi yaitu (61%) sedangkan nilai *recall* *Extreme Gradient Boosting (XGBoost)* lebih rendah yaitu (59%) perbedaan ini memiliki implikasi penting dalam konteks deteksi penyakit diabetes. *Recall* mengukur kemampuan model dalam mengenali semua kasus positif (penderita diabetes). Nilai *recall* yang lebih tinggi pada *Support Vector Machine (SVM)* menunjukkan bahwa model ini lebih mampu mengidentifikasi pasien yang benar-benar menderita diabetes dibandingkan *Extreme Gradient Boosting (XGBoost)*.

Dalam konteks medis, khususnya dalam skrining awal penyakit diabetes, mengurangi jumlah *false negative* (kasus sebenarnya positif tapi diklasifikasikan sebagai negatif) sangat krusial. Kesalahan ini bisa menyebabkan penderita diabetes tidak

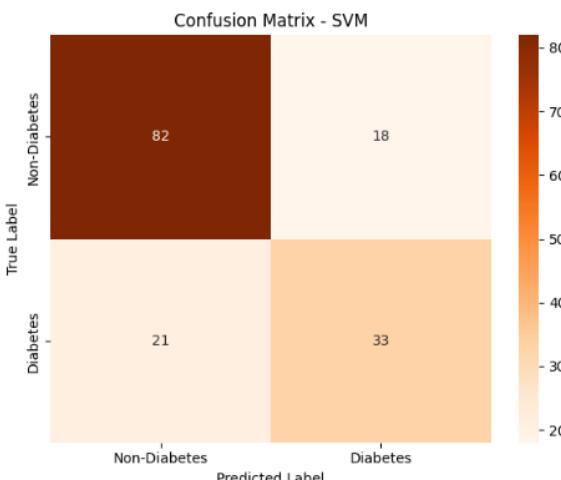
mendapatkan penanganan dini, yang dapat berujung pada komplikasi serius seperti gagal ginjal, gangguan saraf, bahkan risiko kematian. Oleh karena itu, meskipun selisih *recall* hanya 2%, signifikansi klinisnya sangat besar, terutama jika model digunakan dalam sistem pendukung keputusan di layanan kesehatan. Model dengan *recall* lebih tinggi lebih aman dan lebih konservatif, karena cenderung meminimalkan risiko tidak terdeteksinya pasien yang sebenarnya positif.

Model *Hybrid XGBoost-SVM* yang menggabungkan *Extreme Gradient Boosting* (*XGBoost*) dan *Support Vector Machine* (*SVM*) melalui teknik *ensemble voting* memberikan performa terbaik dibanding dua model tunggal. *Akurasi* meningkat menjadi 77%, *presisi* naik menjadi 70%, dan *recall* tetap pada level 61%. *F1-score* yang mencapai 65% menunjukkan bahwa model ini mampu menjaga keseimbangan antara ketepatan dan kemampuan mendeteksi kelas positif dengan lebih baik.

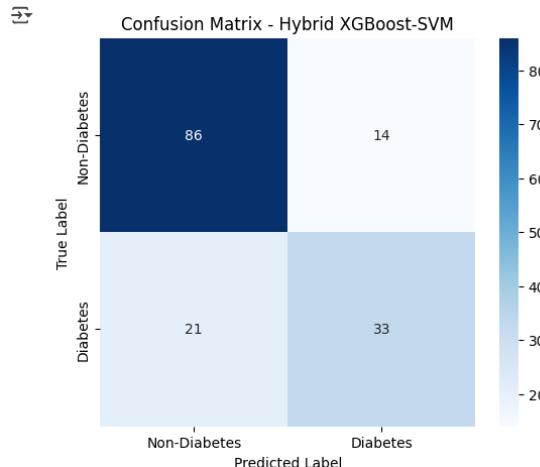
Berdasarkan hasil evaluasi, dapat disimpulkan bahwa model *Hybrid XGBoost-SVM* memberikan hasil yang paling optimal dalam melakukan klasifikasi data diabetes. Pendekatan ensemble ini mampu menggabungkan kelebihan dari masing-masing model dasar sehingga menghasilkan model yang lebih kokoh, akurat, dan seimbang dalam performanya. Dengan demikian, model *Hybrid XGBoost-SVM* ini layak untuk dijadikan sebagai pendekatan utama dalam sistem prediksi serupa yang membutuhkan ketepatan tinggi dan sensitivitas terhadap kelas minoritas.



Gambar 7. Confusion Matrix dari *Extreme Gradient Boosting* (*XGBoost*)



Gambar 8. Confusion Matrix dari *Support Vector Machine* (*SVM*)



Gambar 9. *Confusion Matrix* dari *Hybrid XGBoost-SVM*

Gambar 9. menggambarkan hasil *confusion matrix* dari model *Hybrid XGBoost-SVM*. *Confusion Matrix* ini menggambarkan kinerja model *Hybrid XGBoost-SVM* dalam memprediksi penyakit diabetes. Matrix ini terdiri dari dua baris dan dua kolom, yang merepresentasikan kelas aktual dan kelas prediksi.

1. *True Negatives (TN)*: Terdapat 86 pasien yang sebenarnya tidak menderita diabetes, dan model dengan tepat memprediksi bahwa mereka tidak menderita diabetes.
2. *False Positives (FP)*: Terdapat 14 pasien yang sebenarnya tidak menderita diabetes, namun model salah memprediksi bahwa mereka menderita diabetes.
3. *False Negatives (FN)*: Terdapat 21 pasien yang sebenarnya menderita diabetes, namun model salah memprediksi bahwa mereka tidak menderita diabetes.
4. *True Positives (TP)*: Terdapat 33 pasien yang sebenarnya menderita diabetes, dan model dengan tepat memprediksi bahwa mereka menderita diabetes.

#### 4. KESIMPULAN

Penelitian ini bertujuan untuk membangun sistem pendukung keputusan dalam prediksi penyakit diabetes dengan menerapkan pendekatan *Hybrid XGBoost-SVM* menggunakan algoritma *Extreme Gradient Boosting (XGBoost)* dan *Support Vector Machine (SVM)*. Berdasarkan hasil evaluasi dan analisis yang telah dilakukan, dapat disimpulkan beberapa hal berikut:

1. Model *Hybrid XGBoost-SVM* menunjukkan performa klasifikasi yang lebih baik dibandingkan dengan model tunggal, baik *Extreme Gradient Boosting (XGBoost)* maupun *Support Vector Machine (SVM)*. Model *Hybrid XGBoost-SVM* berhasil mencapai akurasi sebesar 77%, lebih tinggi dibandingkan *Extreme Gradient Boosting (XGBoost)* dan *Support Vector Machine (SVM)* yang masing-masing mencapai 75%.
2. Selain peningkatan akurasi, model *Hybrid XGBoost-SVM* juga menghasilkan presisi lebih tinggi (70%) dan F1-score yang lebih seimbang (65%). Meskipun nilai *recall* model *Hybrid XGBoost-SVM* (61%) setara dengan *Support Vector Machine (SVM)* dan sedikit lebih tinggi dari *Extreme Gradient Boosting (XGBoost)* (59%), perbedaan ini signifikan dalam konteks medis, karena model lebih sensitif dalam mengenali pasien diabetes dan mengurangi risiko *false negative*.
3. Model *Extreme Gradient Boosting (XGBoost)* memiliki keunggulan dalam hal seleksi fitur penting dan efisiensi pemrosesan, sedangkan *Support Vector Machine (SVM)* lebih unggul dalam klasifikasi berbasis margin maksimal, terutama dalam mendeteksi kelas positif. Kombinasi keduanya pada model *Hybrid XGBoost-SVM* memberikan hasil klasifikasi yang lebih stabil dan seimbang.

4. Visualisasi fitur penting yang dihasilkan oleh *Extreme Gradient Boosting (XGBoost)* menunjukkan bahwa fitur *Glucose*, *BMI*, dan *Age* merupakan variabel yang paling berpengaruh dalam menentukan status diabetes seseorang.
5. Berdasarkan evaluasi keseluruhan, pendekatan *Hybrid XGBoost-SVM* layak digunakan sebagai metode utama dalam pengembangan sistem pendukung keputusan untuk deteksi dini penyakit diabetes, karena mampu meningkatkan akurasi, presisi, dan keseimbangan klasifikasi dibanding model individual.

Implementasi metode *Hybrid XGBoost-SVM* ini dapat dijadikan sebagai alternatif yang efektif dalam pengembangan sistem pendukung keputusan di bidang medis, khususnya untuk deteksi dini penyakit diabetes.

## REFERENCES

- Abdurrosyid, R., Teguh, A. and Almais, W. (2025) ‘Deteksi Dini Diabetes menggunakan Machine Learning dengan Metode PCA dan XGBoost’, 11(1), pp. 51–56.
- Agustanti, S.P., Alawiyah, A.P. and Suhandi, N. (2021) ‘Sistem Pendukung Keputusan Diagnosa Penyakit Gigi’, *Sigmata: Jurnal Manajemen dan Informatika*, 9(1), pp. 33–39.
- Amelia, U., Indra, J. and Masruriyah, A.F.N. (2022) ‘Implementasi Algoritma Support Vector Machine (SVM) Untuk Prediksi Penyakit Stroke Dengan Atribut Berpengaruh’, *Scientific Student Journal for Information, Technology and Science*, III(2), pp. 254–259.
- Andryan, M.R., Fajri, M. and Sulistyowati, N. (2022) ‘Komparasi Kinerja Algoritma Xgboost Dan Algoritma Support Vector Machine (SVM) Untuk Diagnosis Penyakit Kanker Payudara’, *JIKO (Jurnal Informatika dan Komputer)*, 6(1), p. 1. Available at: <https://doi.org/10.26798/jiko.v6i1.500>.
- Chang, W. *et al.* (2019) ‘A New Hybrid XGBSVM Model: Application for Hypertensive Heart Disease’, *IEEE Access*, 7, pp. 175248–175258. Available at: <https://doi.org/10.1109/ACCESS.2019.2957367>.
- Debora Mait, C. *et al.* (2022) ‘Sistem Pendukung Keputusan Menggunakan Fuzzy Logic Tahani Untuk Penentuan Golongan Obat Sesuai Dengan’, *Jurnal Media Infotama*, 18(2), pp. 344–353.
- Erlin *et al.* (2022) ‘Deteksi Dini Penyakit Diabetes Menggunakan Machine Learning dengan Algoritma Logistic Regression’, *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, 11(2), pp. 88–96. Available at: <https://doi.org/10.22146/jnteti.v11i2.3586>.
- Karyadiputra, E. and Setiawan, A. (2023) ‘Sistem pendukung keputusan berbasis decision tree algorithm untuk prediksi penyakit diabetes’, *Media Informasi dan Teknologi*, 17(2023), pp. 294–301. Available at: <https://doi.org/10.24252/teknosains.v17i3.38383>.
- Khairunnisa, A. (2023) ‘Perbandingan Model Random Forest Dan Xgboost Untuk Prediksi Kejahatan Kesusilaan Di Provinsi Jawa Barat’, *JIKO (Jurnal Informatika dan Komputer)*, 7(2), p. 202. Available at: <https://doi.org/10.26798/jiko.v7i2.799>.
- Kurnia, D. *et al.* (2023) ‘Seleksi Fitur dengan Particle Swarm Optimization pada Klasifikasi Penyakit Parkinson Menggunakan XGBoost’, *Jurnal Teknologi Informasi dan Ilmu Komputer*, 10(5), pp. 1083–1094. Available at: <https://doi.org/10.25126/jtiik.20231057252>.
- Lestari, U.I., Nadhiroh, A.Y. and Novia, C. (2021) ‘Penerapan Metode K-Nearest Neighbor Untuk Sistem Pendukung Keputusan Identifikasi Penyakit Diabetes Melitus’, 8(4), pp. 2071–2082.
- Mucholladin, A.W., Bachtiar, F.A. and Furqon, M.T. (2021) ‘Klasifikasi Penyakit Diabetes menggunakan Metode Support Vector Machine’, *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 5(2), pp. 622–633. Available at: <http://j-ptiik.ub.ac.id>.

- Nasution, M.K., Saedudin, R.R. and Widartha, V.P. (2021) 'Perbandingan Akurasi Algoritma Naïve Bayes Dan Algoritma Xgboost Pada Klasifikasi Penyakit Diabetes', *e-Proceeding of Engineering*, 8(5), pp. 9765–9772. Available at: <https://journal.ubpkarawang.ac.id/mahasiswa/index.php/ssj/article/view/424/338> <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/15759>.
- Oktavia, D.C., Aeni, K. and Saraswati, N.M. (2020) 'Sistem Pendukung Keputusan Menu Makanan Untuk Penderita Penyakit Tipes Dan Diabetes Menggunakan Metode Technique For Order Preference By Similarity To Ideal Solution (TOPSIS) (Studi Kasus : RSUM SA)', *IJIR ( Indonesian Journal of Informatics and Research)*, 1(1), pp. 8–13.
- Ramadhanti, A. *et al.* (2024) 'Pengembangan Alat Deteksi Gejala Anemia Non-Invasive Berbasis Narrowband Internet of ', 12(3).
- Rayadin, M.A. *et al.* (2024) 'Implementasi Ensemble Learning Metode XGBoost dan Random Forest untuk Prediksi Waktu Penggantian Baterai Aki', *BIOS: Jurnal Teknologi Informasi dan Rekayasa Komputer*, 5(2), pp. 111–119.
- Soelistijadi, R. *et al.* (2024) 'Pemodelan Prediktif Menggunakan Metode Ensemble Learning XGBoost dalam Peningkatan Akurasi Klasifikasi Penyakit Ginjal', 5(4), pp. 1866–1875.
- Syukron, M., Santoso, R. and Widiharih, T. (2020) 'Perbandingan Metode Smote Random Forest Dan Smote Xgboost Untuk Klasifikasi Tingkat Penyakit Hepatitis C Pada Imbalance Class Data', *Jurnal Gaussian*, 9(3), pp. 227–236. Available at: <https://doi.org/10.14710/j.gauss.v9i3.28915>.
- Wahyuni, D. (2025) 'Jurnal Sains dan Teknologi Widyaloka Sistem Pendukung Keputusan Dengan Metode Weighted Product Untuk Memilih Protokol Pengobatan Pneumonia Pada Anak Berdasarkan Kondisi Klinis Jurnal Sains dan Teknologi Widyaloka', 4, pp. 70–75.
- Wibowo, Indarti and Laraswati, D. (2024) 'Komparasi Algoritma Decision Tree, Random Forest dan SVM untuk Prognosis COVID-19', *IMTechno: Journal of Industrial Management and Technology*, 5(2), pp. 10–15. Available at: <https://www.kaggle.com/datasets/maulanazhahran/d>.