

# Machine Learning Analysis for Diabetes Prediction

Decision Tree vs SVM  
Model Comparison

Nadhif Rif'at Rasendriya

# Overview

Proyek ini bertujuan untuk memprediksi apakah seseorang menderita diabetes atau tidak berdasarkan beberapa fitur kesehatan, seperti usia, BMI, tekanan darah, dan kadar glukosa.

Model machine learning yang digunakan adalah:

- Decision Tree 🌳
- Support Vector Machine (SVM) ⚡

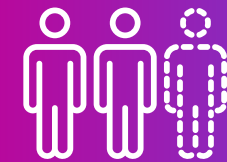
# Dataset yang digunakan

- ✓ Data Training → Digunakan untuk melatih model
- ✓ Data Testing → Digunakan untuk mengevaluasi model

## Fitur Dalam Dataset



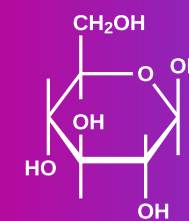
Age (Usia) → Umur pasien



BMI (Body Mass Index) → Indeks massa tubuh



Blood Pressure → Tekanan darah pasien



Glucose (Kadar Glukosa) →  
Jumlah glukosa dalam darah



Diabetes (Label) →  
1 jika pasien menderita diabetes, 0 jika tidak

# Exploratory Data Analysis (EDA)

Sebelum membangun model, dilakukan analisis eksplorasi data (EDA) untuk memahami pola dalam dataset. Beberapa langkah yang dilakukan dalam EDA adalah:



## a. Memeriksa Distribusi Data

- Melihat bagaimana data seperti usia, BMI, tekanan darah, dan glukosa tersebar.
- Menganalisis apakah ada outlier atau nilai ekstrim yang dapat mempengaruhi model.



## b. Korelasi Antar Fitur

- Menganalisis hubungan antara BMI, tekanan darah, dan glukosa terhadap diabetes.
- Jika suatu fitur memiliki hubungan kuat dengan diabetes, maka fitur tersebut akan berpengaruh besar pada prediksi.



## c. Data Cleaning

- Memeriksa missing values atau data yang hilang.
- Normalisasi atau standarisasi data jika diperlukan.

# Full Code

[https://github.com/nadhif-royal/DecisionTree\\_vs\\_SVM\\_DiabetesPrediction](https://github.com/nadhif-royal/DecisionTree_vs_SVM_DiabetesPrediction)



Scan here



# Datasets

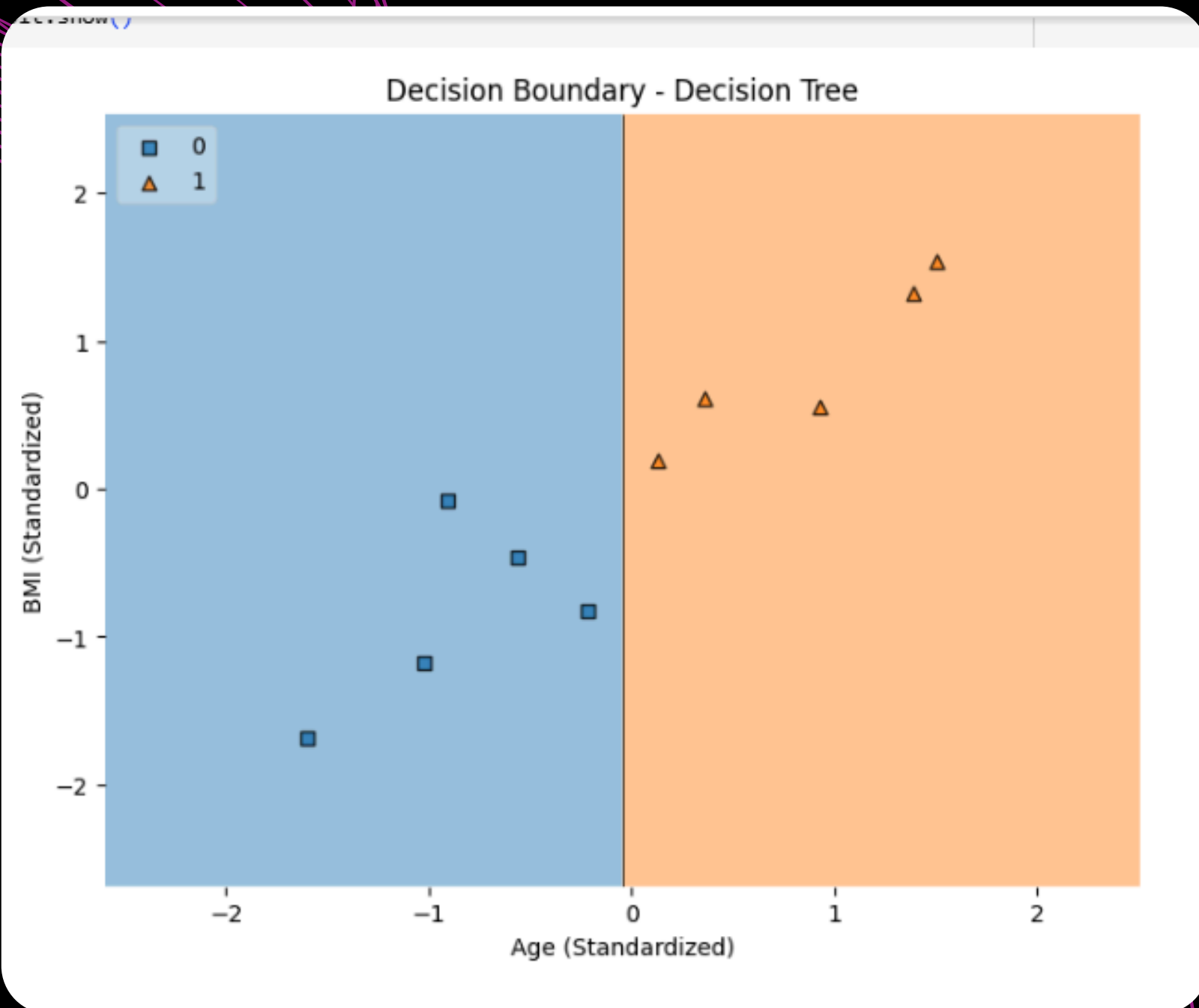
## Train

	Age	BMI	BloodPressure	Glucose	Diabetes
0	45	25.3	120	90	0
1	50	30.1	140	160	1
2	39	27.8	130	105	0
3	60	33.2	145	180	1
4	33	22.4	110	85	0
5	55	29.9	135	150	1
6	42	26.5	125	95	0
7	48	28.7	138	145	1
8	59	32.5	142	175	1
9	38	24.1	118	100	0

## Test

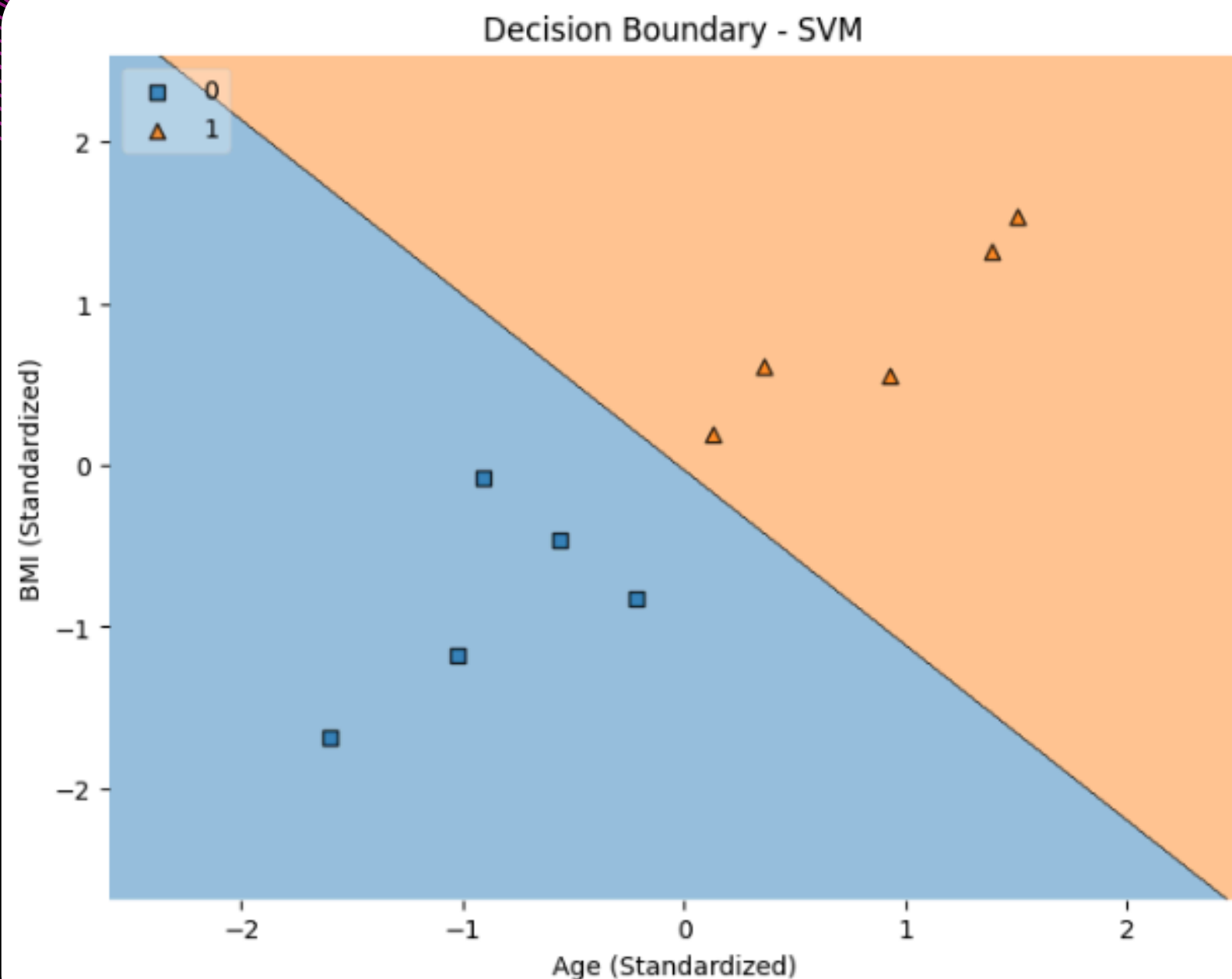
	Age	BMI	BloodPressure	Glucose	Diabetes
0	52	29.5	137	155	1
1	41	25.7	122	98	0
2	36	22.9	115	88	0
3	58	31.2	140	170	1
4	47	27.3	128	110	0

# Visualization Decision Tree



- Model Decision Tree membagi ruang fitur dengan garis tegas (vertikal) yang menunjukkan batas keputusan yang tidak terlalu kompleks.
- Model tampaknya memisahkan kelas berdasarkan nilai Age, dengan pemisahan utama di sekitar Age = 0 (standar).
- Hal ini menunjukkan bahwa Decision Tree lebih cenderung membuat keputusan berbasis aturan sederhana tanpa mempertimbangkan pola yang lebih kompleks dalam data.

# Visualization SVM (Support Vector Machine)



Grafik ini menampilkan Decision Boundary (Batas Keputusan) dari model SVM (Support Vector Machine) yang digunakan untuk membedakan dua kelas berdasarkan Age (Standardized) dan BMI (Standardized).

- SVM pada grafik ini menunjukkan keputusan linier yang cukup baik dalam membagi dua kelas berdasarkan fitur Age dan BMI.
- Batas keputusan diagonal menunjukkan bahwa kedua fitur memiliki kontribusi dalam pemisahan kelas.
- Jika dataset lebih kompleks, bisa mempertimbangkan kernel SVM agar hasilnya lebih optimal.



# The Result

## Decision Tree

- ✓ Akurasi: 0.80
- ✓ F1 Score: 0.80
- ✓ Mean Squared Error: 0.20
- ✓  $R^2$  Score: 0.17

## SVM

- ✓ Akurasi: 1.00
- ✓ F1 Score: 1.00
- ✓ Mean Squared Error: 0.00
- ✓  $R^2$  Score: 1.00

# Perbandingan Model Decision Tree vs SVM

Berdasarkan hasil evaluasi, berikut adalah perbandingan antara Decision Tree dan Support Vector Machine (SVM) dalam memprediksi diabetes:

Metode	Akurasi	F1 Score	MSE	R <sup>2</sup> Score
Decision Tree	0.80	0.80	0.20	0.17
SVM	1.00	1.00	0.00	1.00

# Analisis Perbandingan

## Akurasi & F1 Score

- SVM memiliki akurasi dan F1 Score 1.00 (100%), yang berarti model ini mampu mengklasifikasikan semua data dengan sempurna tanpa kesalahan.
- Decision Tree hanya mencapai akurasi 80% dan F1 Score 0.80, yang berarti ada beberapa kesalahan prediksi.

## Mean Squared Error (MSE)

- SVM memiliki MSE 0.00, yang berarti tidak ada kesalahan dalam prediksi.
- Decision Tree memiliki MSE 0.20, yang menunjukkan masih ada tingkat kesalahan dalam model ini.

## R<sup>2</sup> Score

- SVM memiliki nilai R<sup>2</sup> Score = 1.00, yang berarti model ini mampu menjelaskan 100% variabilitas dalam data.
- Decision Tree memiliki R<sup>2</sup> Score = 0.17, yang cukup rendah, menunjukkan bahwa model ini kurang baik dalam menjelaskan variasi data.

# Kesimpulan

- ◆ SVM adalah model terbaik untuk dataset ini, karena memberikan hasil sempurna dengan akurasi, F1 Score, dan  $R^2$  Score sebesar 1.00, serta tidak memiliki kesalahan prediksi ( $MSE = 0.00$ ).
- ◆ Decision Tree masih cukup baik dengan akurasi 80%, tetapi performanya lebih rendah dibandingkan SVM. Model ini bisa menjadi pilihan jika ingin interpretasi lebih mudah, tetapi kurang optimal dalam kasus ini.



# Terimakasih



[linkedin.com/in/royalnadhif50/](https://www.linkedin.com/in/royalnadhif50/)