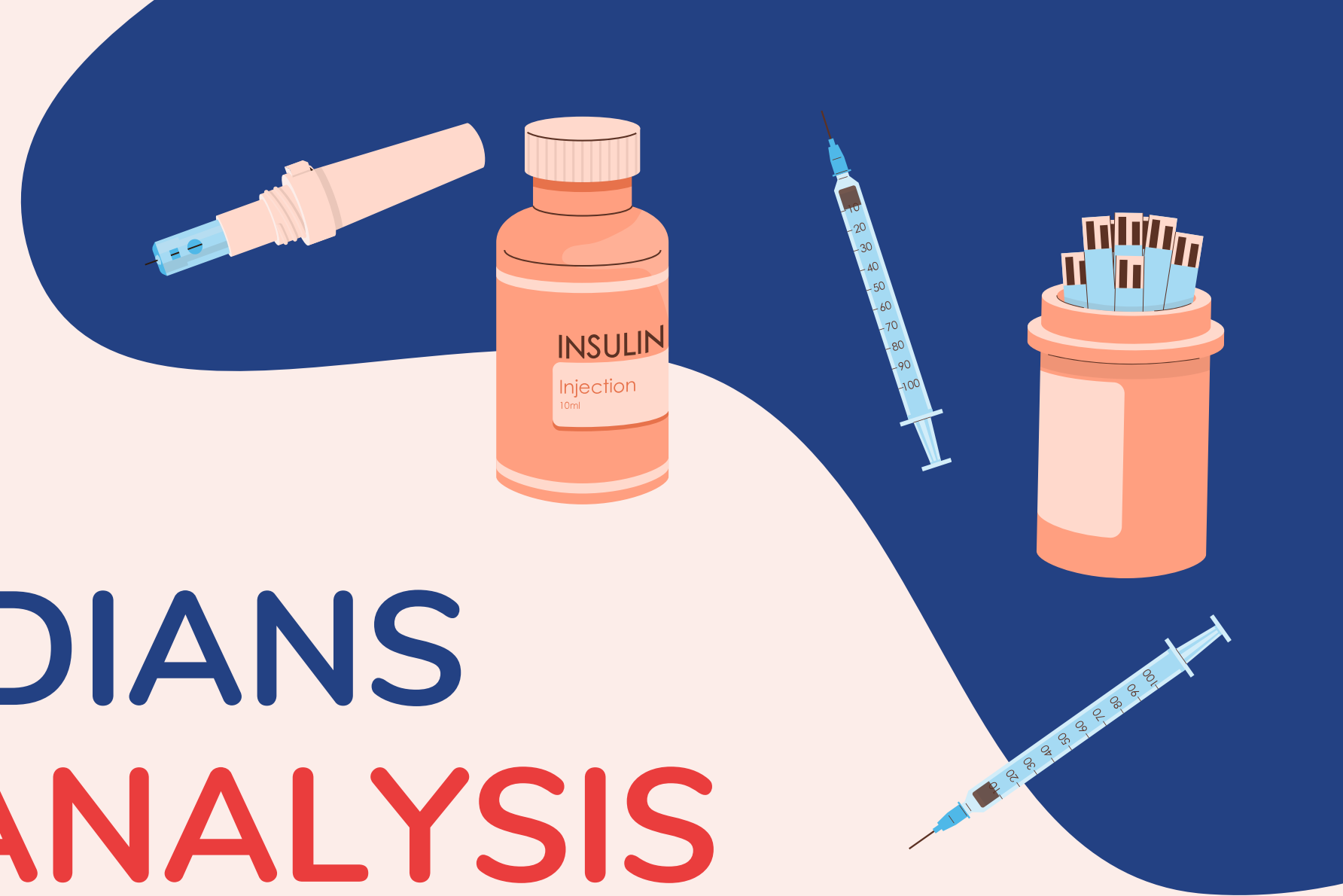


PIMA INDIANS DIABETES ANALYSIS

USING KNN AND K-MEANS





DAFTAR ISI

- About Data
- Tujuan Analisis
- Feature Selection
- K-Means
- KNN
- Perbandingan KNN
- Kesimpulan
- Saran Bisnis

ABOUT DATA

Data pasien yang menderita Diabetes atau tidak berdasarkan sejumlah pengukuran diagnostik yang tercantum dalam dataset.

Seluruh pasien dalam dataset ini adalah wanita berusia minimal 21 tahun dengan keturunan Suku Indian Pima.

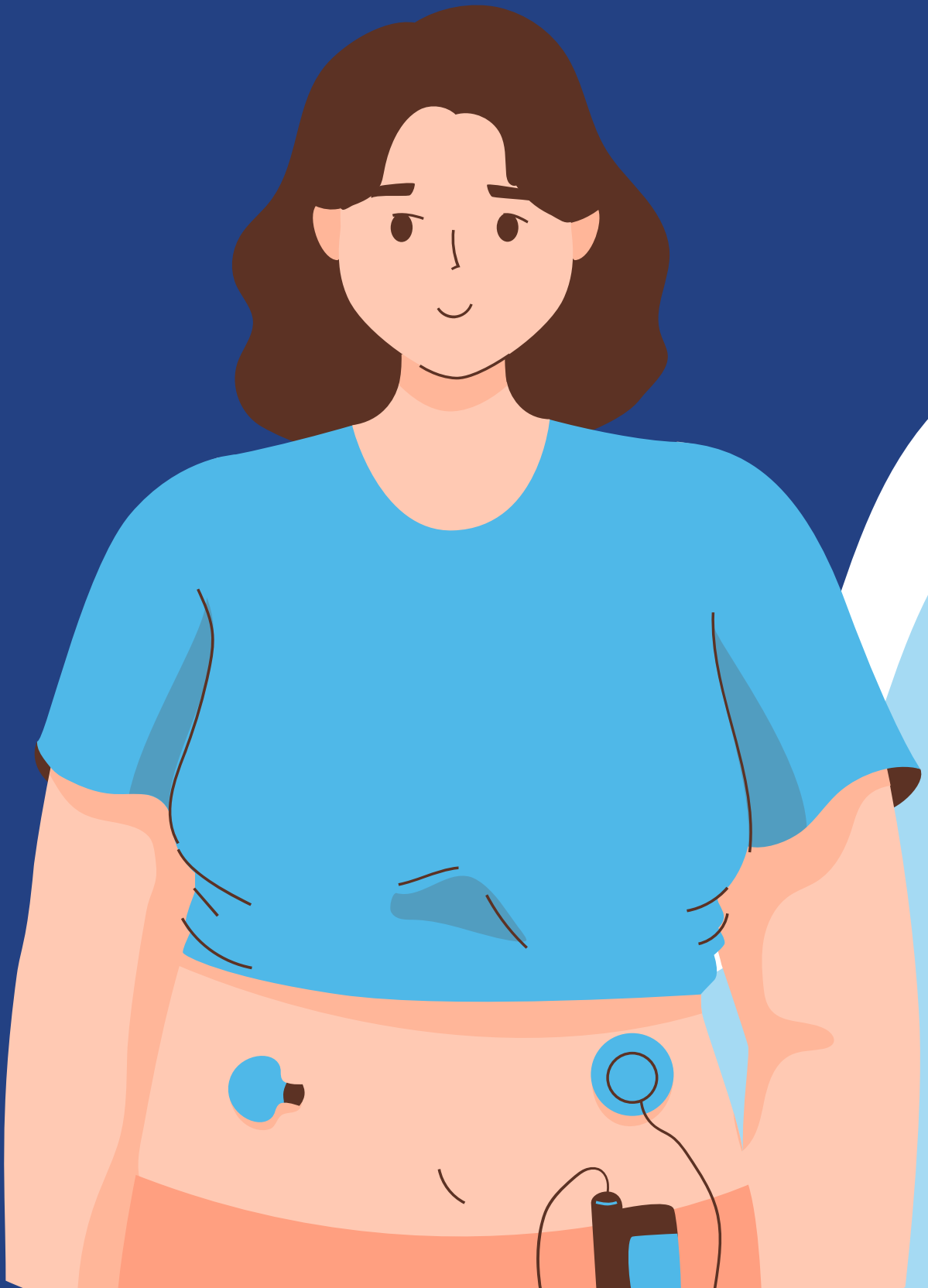
- Pregnancies
- Glucose
- Blood Pressure
- Skin Thickness
- Insulin
- BMI
- Diabetes Pedigree Function
- Age
- Outcome

 **768** Row Data

Pima_Indians_Diabetes_with_header



PREGNANCIES	Jumlah kehamilan yang pernah dialami oleh pasien.
GLUCOSE	Kadar glukosa dalam darah, nilai tinggi dapat mengindikasikan prediabetes atau diabetes.
BLOOD PRESSURE	Tekanan darah yang tinggi seringkali berhubungan dengan risiko komplikasi diabetes.
SKIN THICKNESS	Ketebalan lipatan kulit triceps dalam mm, digunakan untuk mengukur kadar lemak tubuh.
INSULIN	Ketidakseimbangan insulin menjadi indikator kuat untuk diabetes.
BMI	Body Mass Index, yaitu berat badan (kg) dibagi tinggi badan kuadrat (m ²).
DIABETES PEDIGREE FUNCTION	Fungsi silsilah keluarga yang mengukur kemungkinan terkena diabetes berdasarkan riwayat genetik.
AGE	Usia pasien dalam tahun.
OUTCOME	Label target: 0 = Tidak menderita diabetes 1 = Menderita diabetes. Ini merupakan hasil diagnosis berdasarkan semua parameter di atas.



TUJUAN ANALISIS

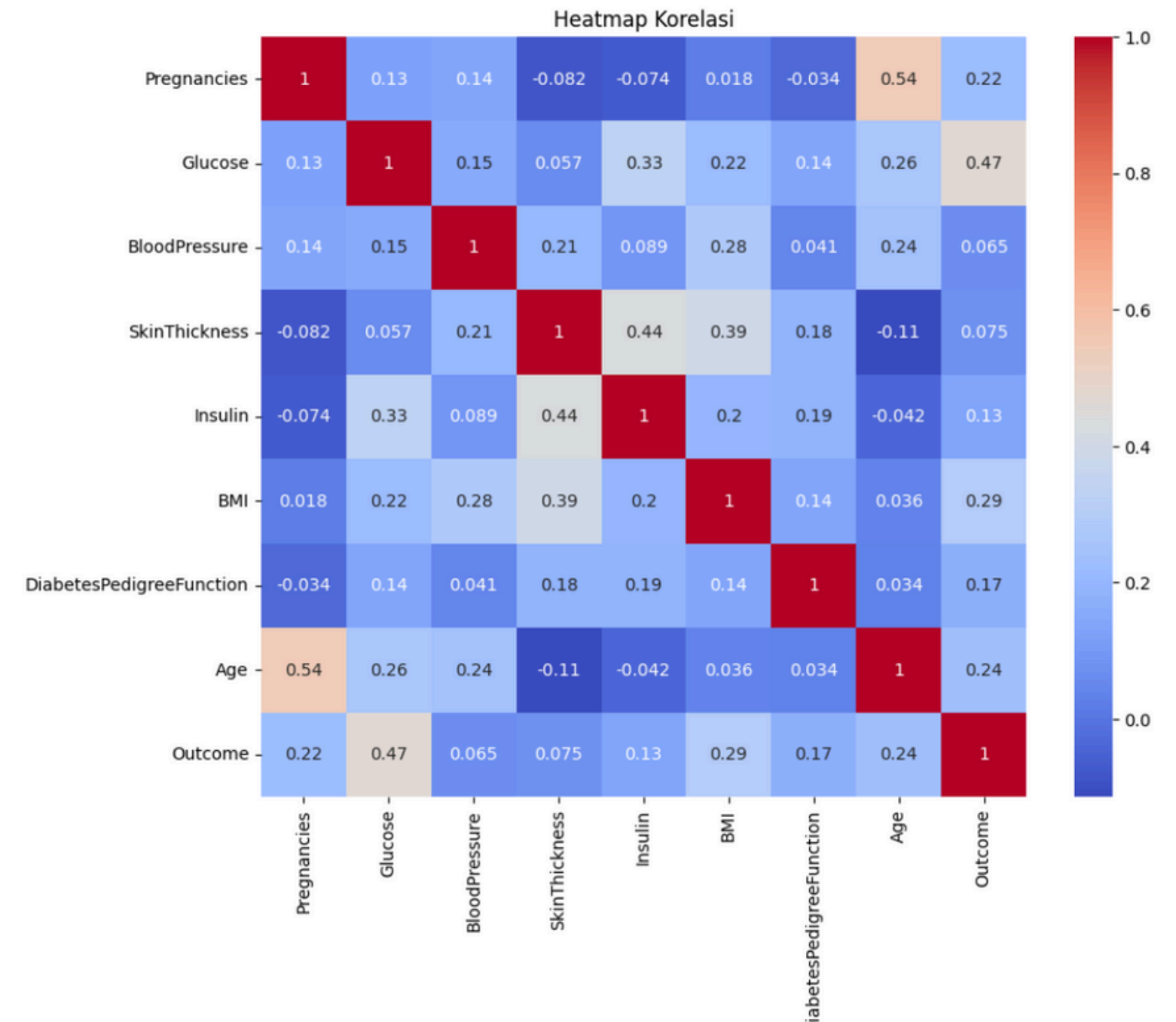
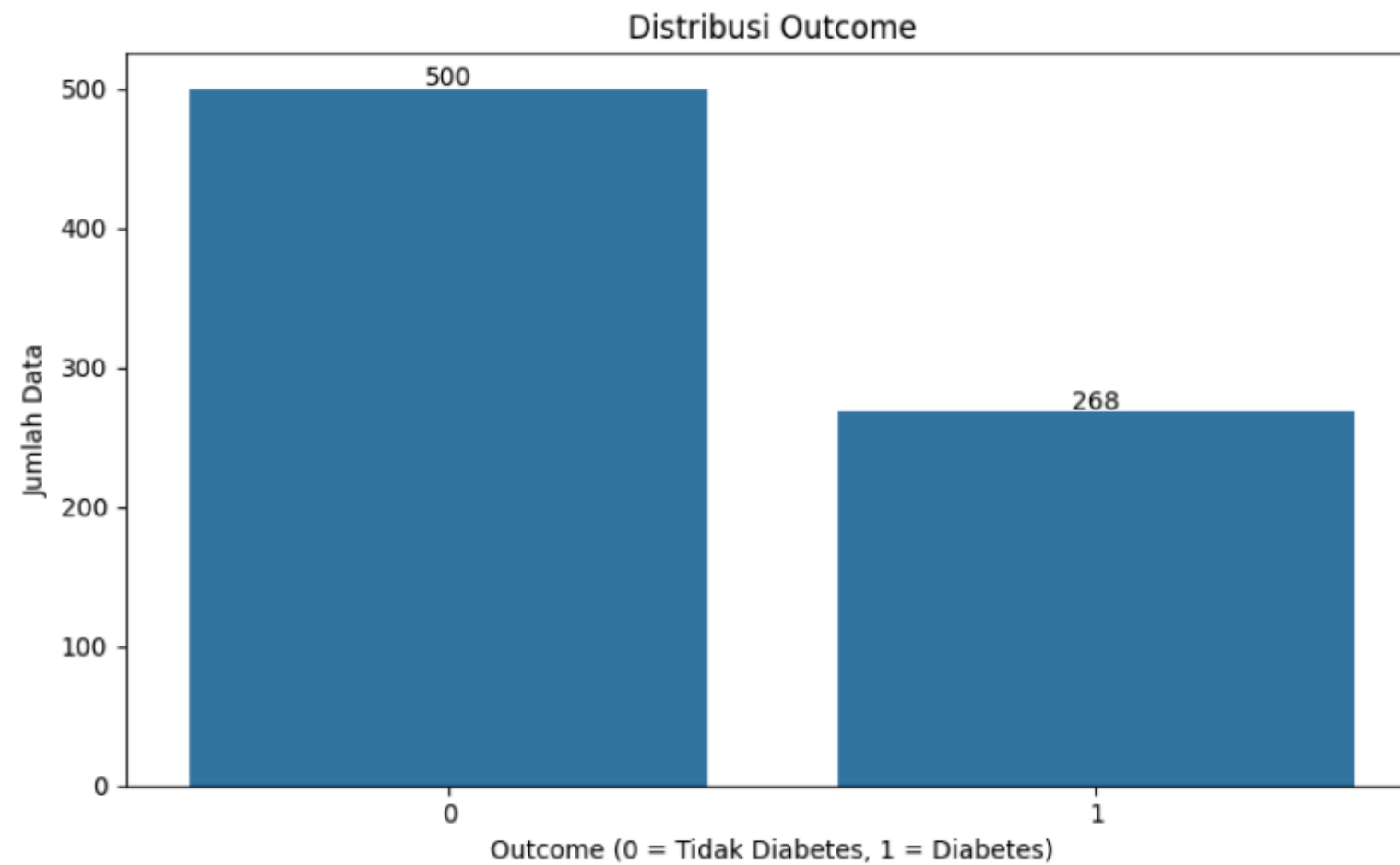


Menerapkan metode K-Nearest Neighbors (KNN) dan K-Means untuk mengcluster pasien diabetes.

Membandingkan hasil KNN menggunakan Outcome Actual dengan hasil KNN menggunakan Outcome K-Means

Mengevaluasi Segmentasi Pasien Menggunakan KMeans Clustering

FEATURE SELECTION

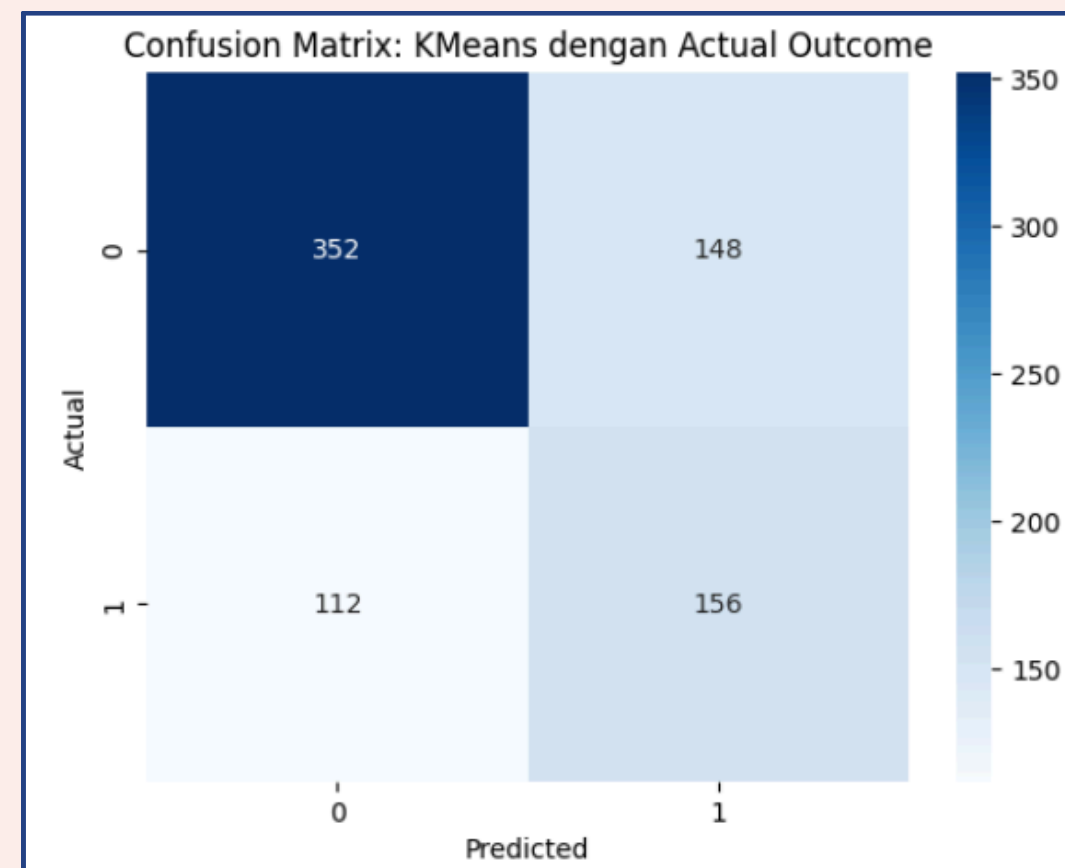


- Tidak ada variabel yang berkorelasi negatif kuat dengan Outcome
- Karena KNN mempertimbangkan semua fitur dalam penghitungan jarak, maka semakin banyak informasi, semakin baik akurasi jika fitur **tidak redundant**.

K-Means

2 Cluster

Metode **unsupervised learning** (pembelajaran tanpa label) yang digunakan untuk mengelompokkan data ke dalam beberapa grup (cluster) berdasarkan kemiripan fitur.



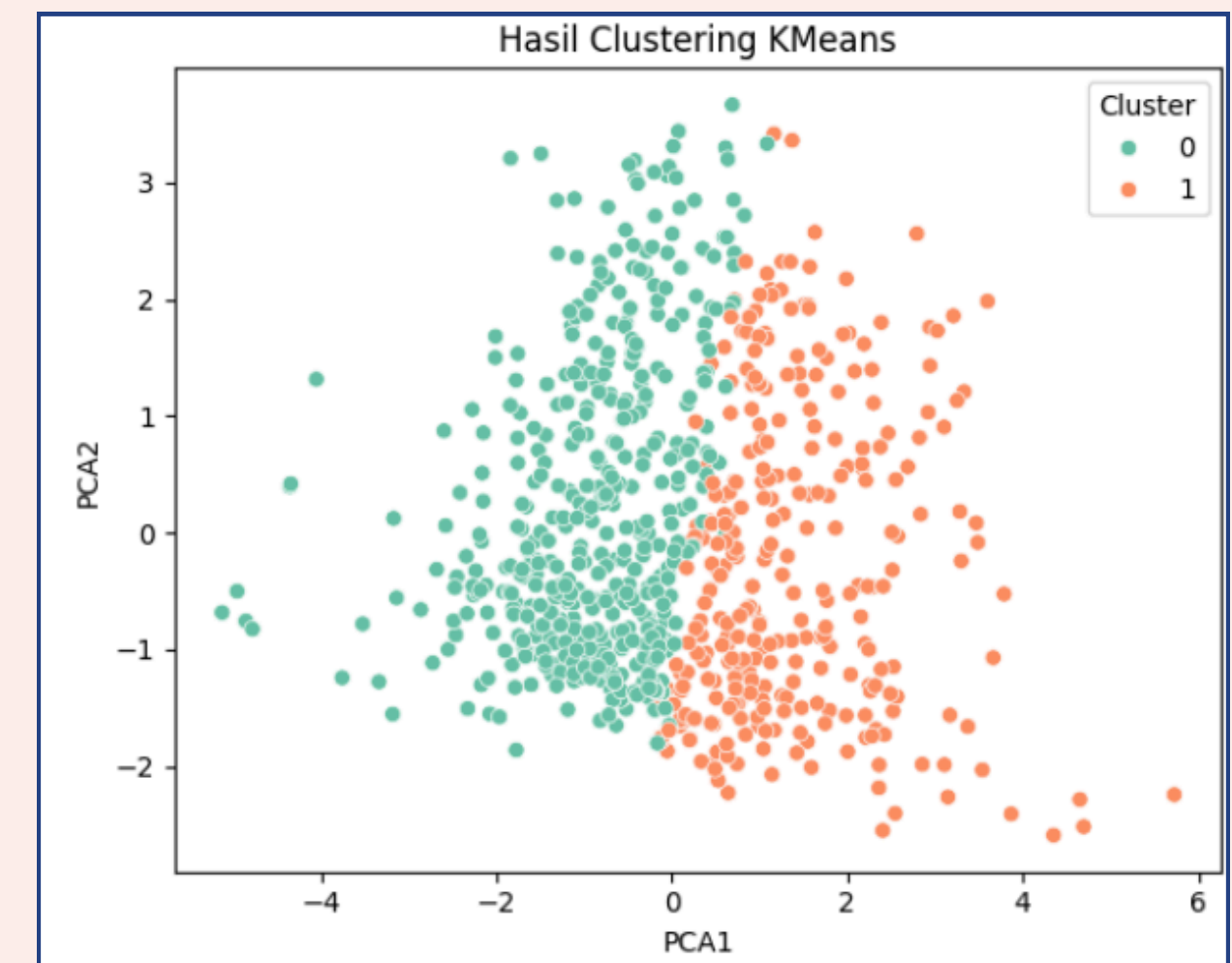
=== Evaluasi K-Means Clustering ===

Accuracy : 0.6614583333333334

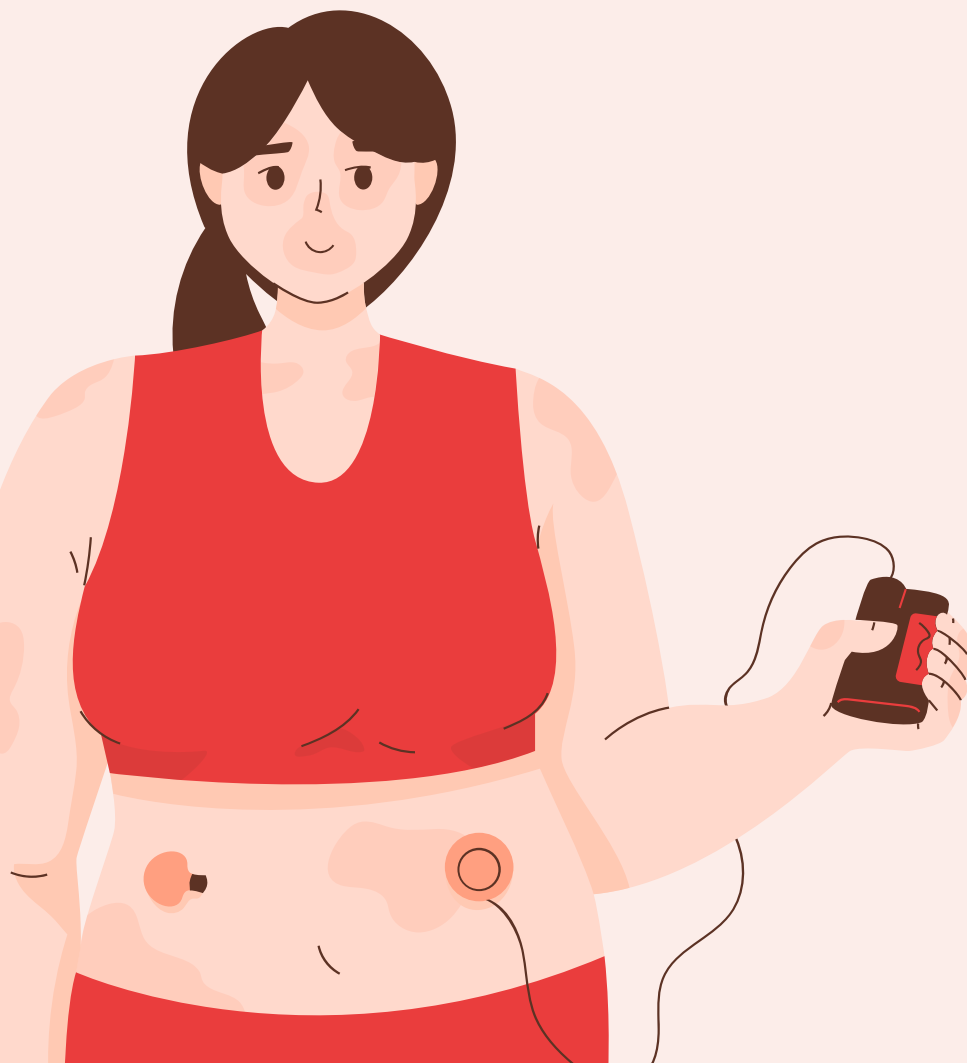
Precision: 0.5131578947368421

Recall : 0.582089552238806

F1 Score : 0.5454545454545454

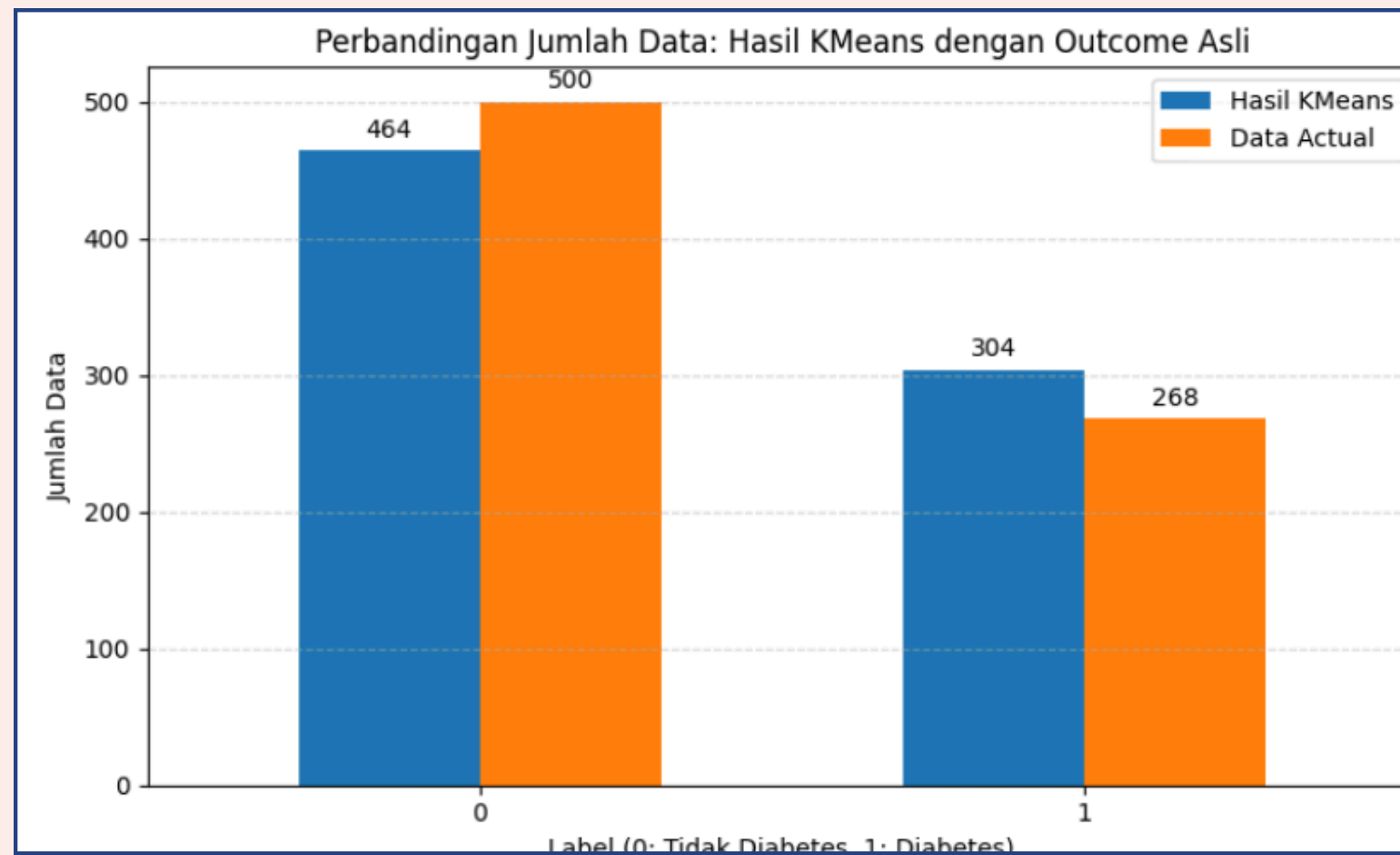


- Kedua cluster terlihat cukup terpisah di sepanjang sumbu PCA1, meskipun terdapat sedikit area tumpang tindih di bagian tengah.
- Ini menunjukkan bahwa KMeans berhasil membentuk dua cluster utama berdasarkan kemiripan pola data.



K-Means

2 Cluster



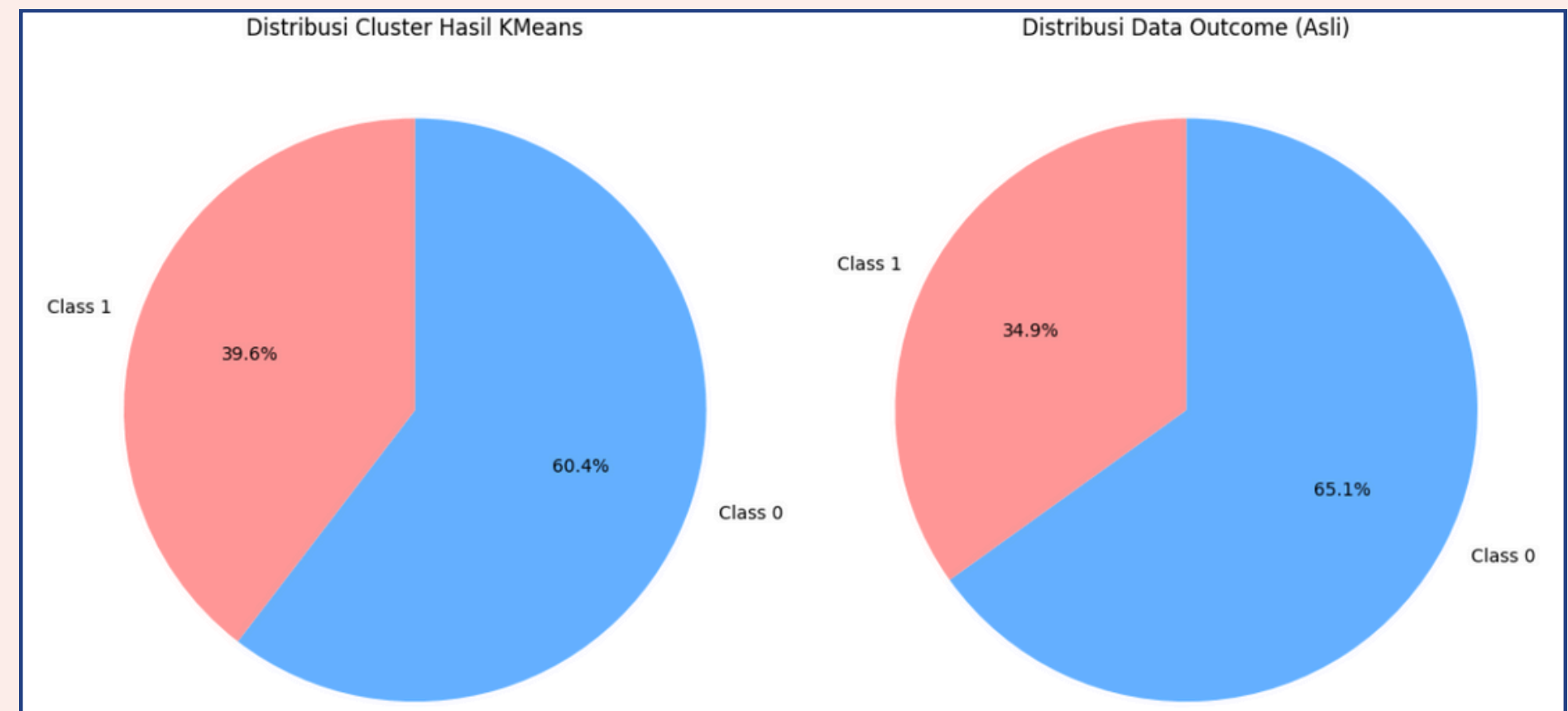
1. Distribusi Hasil KMeans

- Cluster 0: **60.4%**
- Cluster 1: **39.6%**

2. Distribusi Outcome Asli

- Outcome 0 (Tidak Diabetes): **65.1%**
- Outcome 1 (Diabetes): **34.9%**

- Hasil clustering relatif sejalan dengan distribusi actual pada dataset.
- Ini menunjukkan bahwa KMeans cukup mampu memisahkan pasien ke dalam dua kelompok risiko tanpa mengetahui label asli.



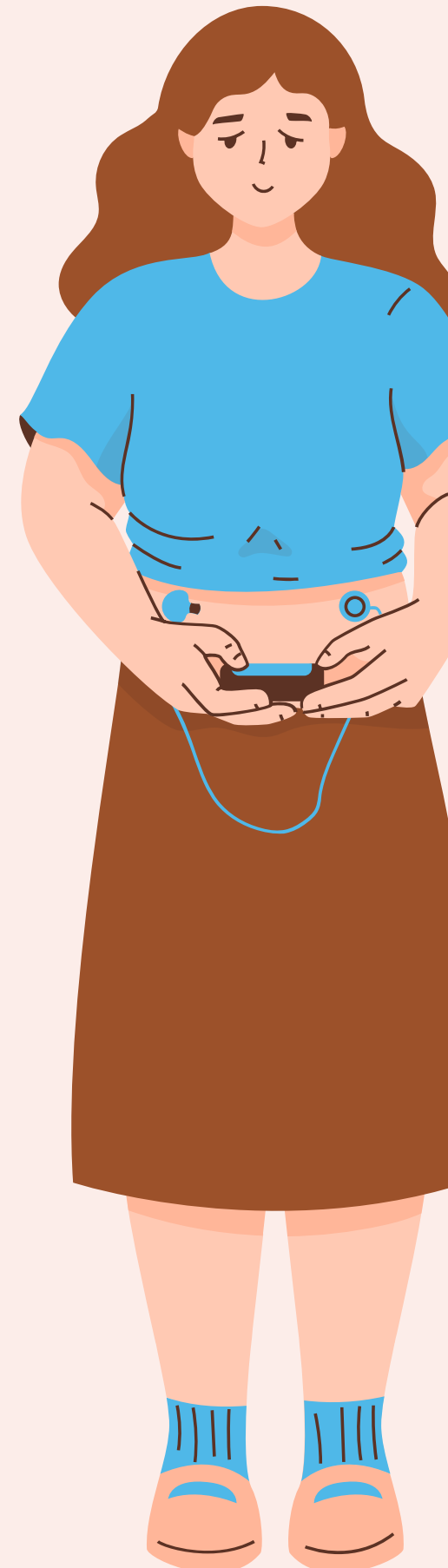
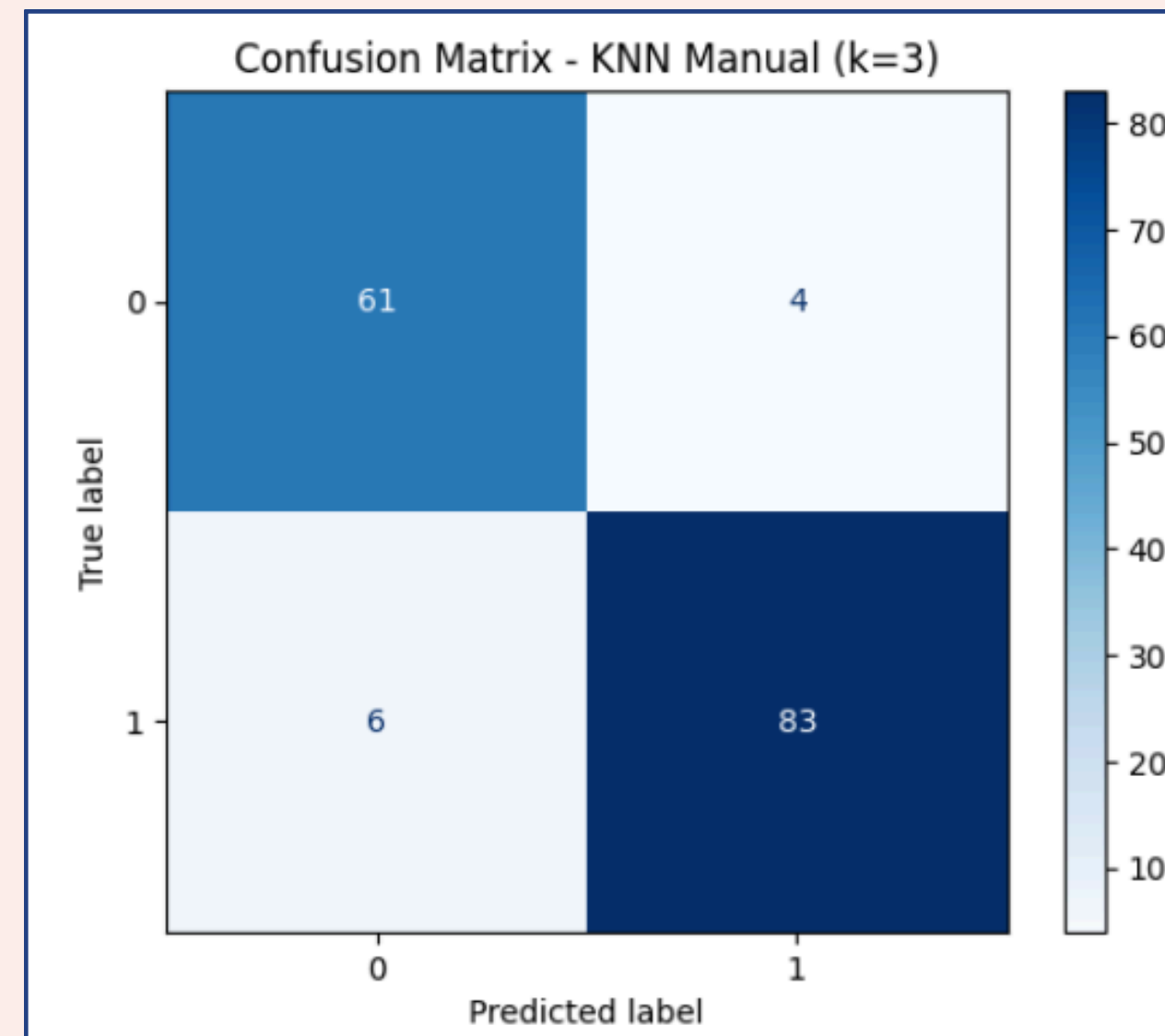
KNN

dengan Output K-Means

KNN adalah metode machine learning **supervised classification** yang memprediksi label suatu data baru berdasarkan kemiripannya (jarak terdekat) dengan data-data yang sudah diketahui labelnya.

🔍 Evaluasi Akurasi untuk Berbagai Nilai k:

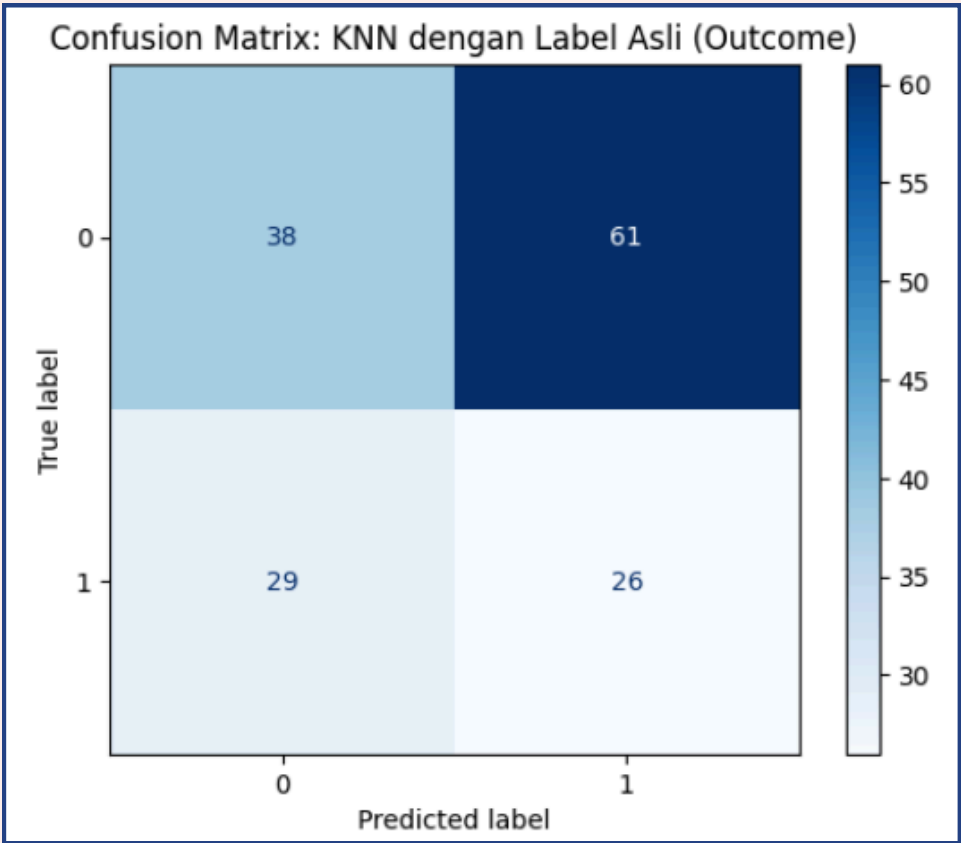
- k = 1 → Akurasi = 92.86%
- k = 3 → Akurasi = 93.51%**
- k = 5 → Akurasi = 91.56%
- k = 7 → Akurasi = 90.91%
- k = 9 → Akurasi = 90.26%
- k = 11 → Akurasi = 90.91%



Perbandingan KNN

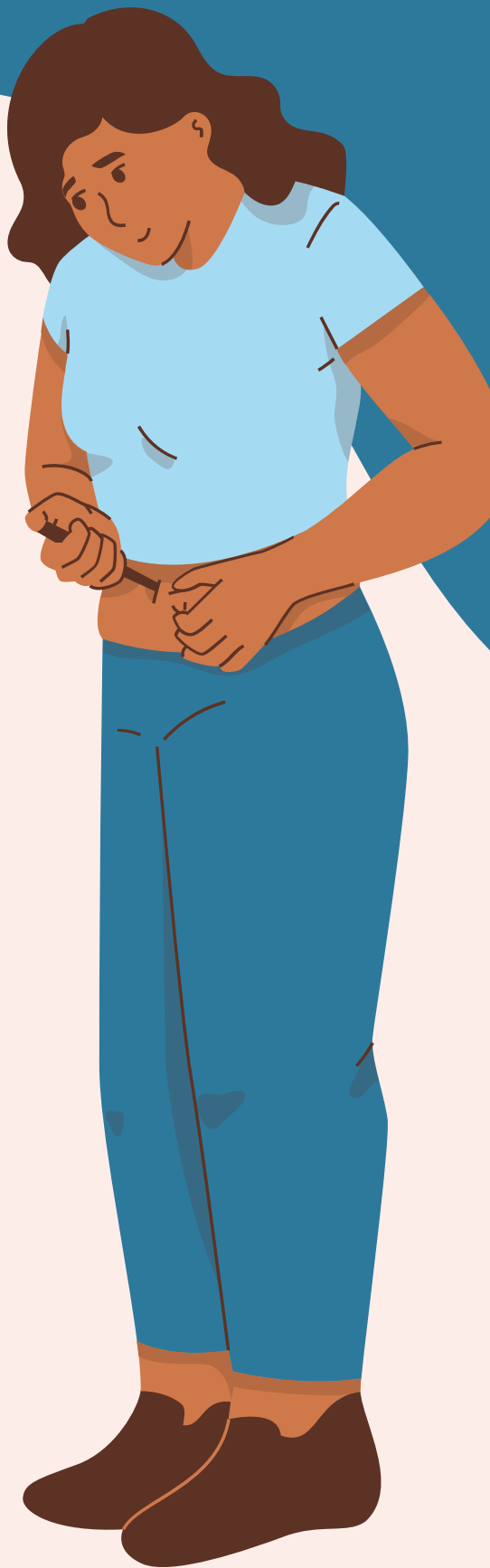
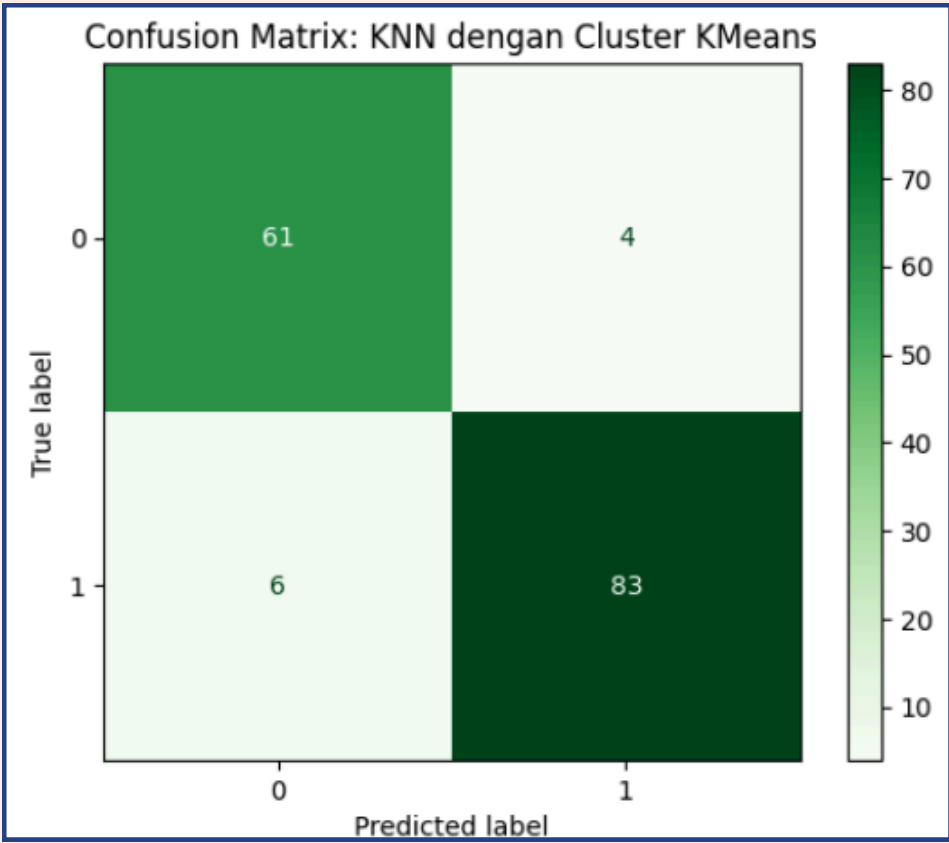
dengan Label Asli (Outcome)

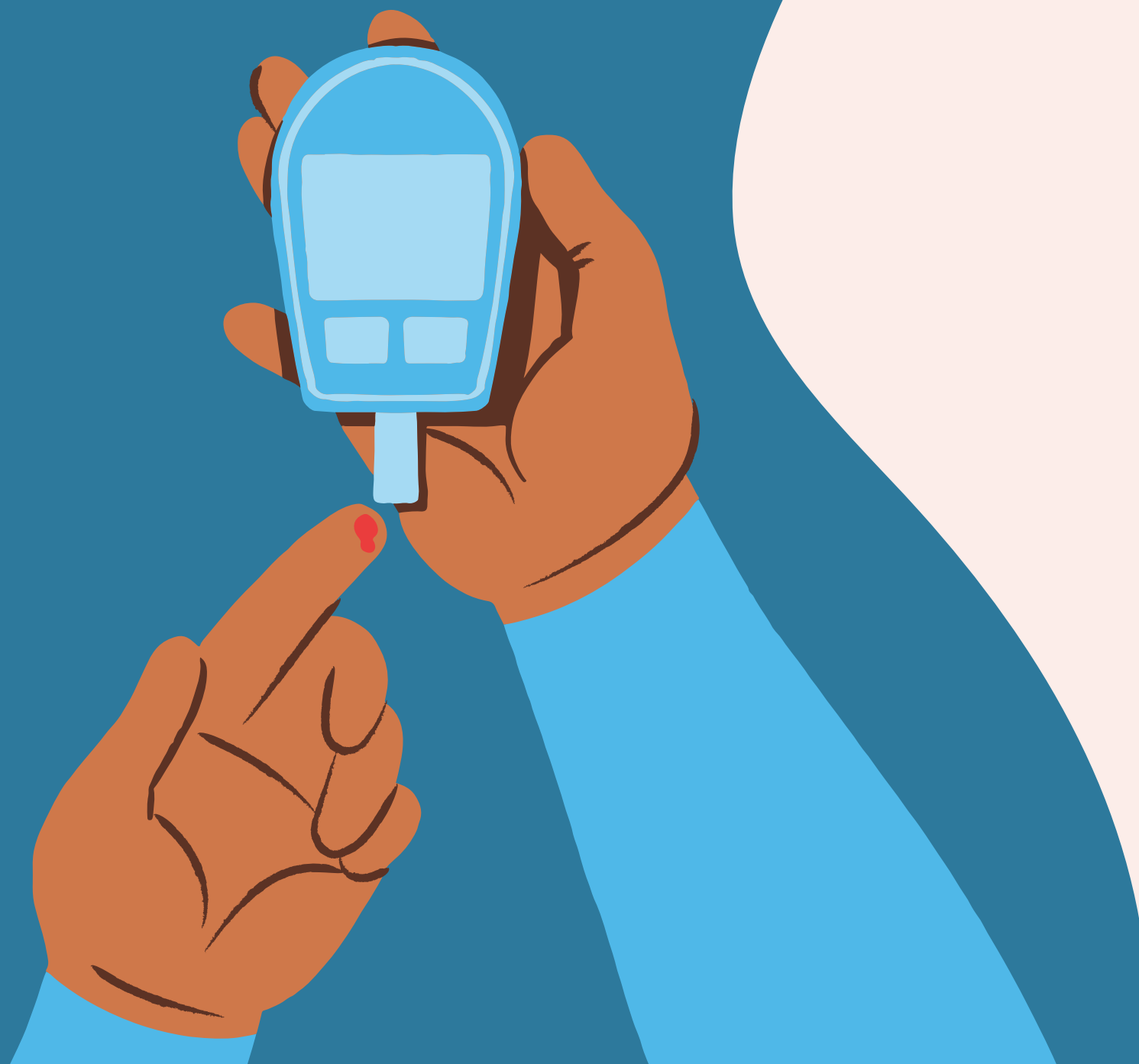
Evaluasi KNN terhadap Label Asli (Outcome):				
Akurasi : 41.56%				
Precision: 29.89%				
Recall : 47.27%				
F1-Score : 36.62%				
Classification Report:				
	precision	recall	f1-score	support
0	0.57	0.38	0.46	99
1	0.30	0.47	0.37	55
accuracy			0.42	154
macro avg	0.43	0.43	0.41	154
weighted avg	0.47	0.42	0.43	154



dengan Output K-Means

Evaluasi KNN terhadap Cluster dari KMeans:				
Akurasi : 93.51%				
Precision: 95.40%				
Recall : 93.26%				
F1-Score : 94.32%				
Classification Report (vs Cluster KMeans):				
	precision	recall	f1-score	support
0	0.91	0.94	0.92	65
1	0.95	0.93	0.94	89
accuracy			0.94	154
macro avg	0.93	0.94	0.93	154
weighted avg	0.94	0.94	0.94	154





Cluster_Outcome	0	1	All
Outcome			
0	352	148	500
1	112	156	268
All	464	304	768

- Cluster 0 berhasil mengelompokkan 352 dari 500 pasien yang tidak diabetes.
- Cluster 1 berhasil mengelompokkan 156 dari 268 pasien yang diabetes.

Artinya, clustering berhasil mengidentifikasi pola dasar dalam dataset.

KESIMPULAN

- KMeans berhasil membentuk dua cluster utama (0 dan 1) yang mencerminkan distribusi asli pasien diabetes dan non-diabetes.
- Model KNN berhasil memberikan akselerasi klasifikasi dengan hasil terbaik pada $k = 3$ dengan Akurasi: 93.51%
- KNN juga diuji dengan target label dari hasil KMeans, dan memberikan performa tinggi :

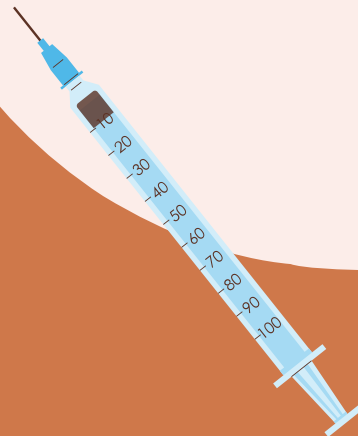
Akurasi: 93.51%

Ini menunjukkan bahwa pola cluster dari KMeans dapat direplikasi secara akurat oleh KNN, menjadikannya model klasifikasi yang andal.

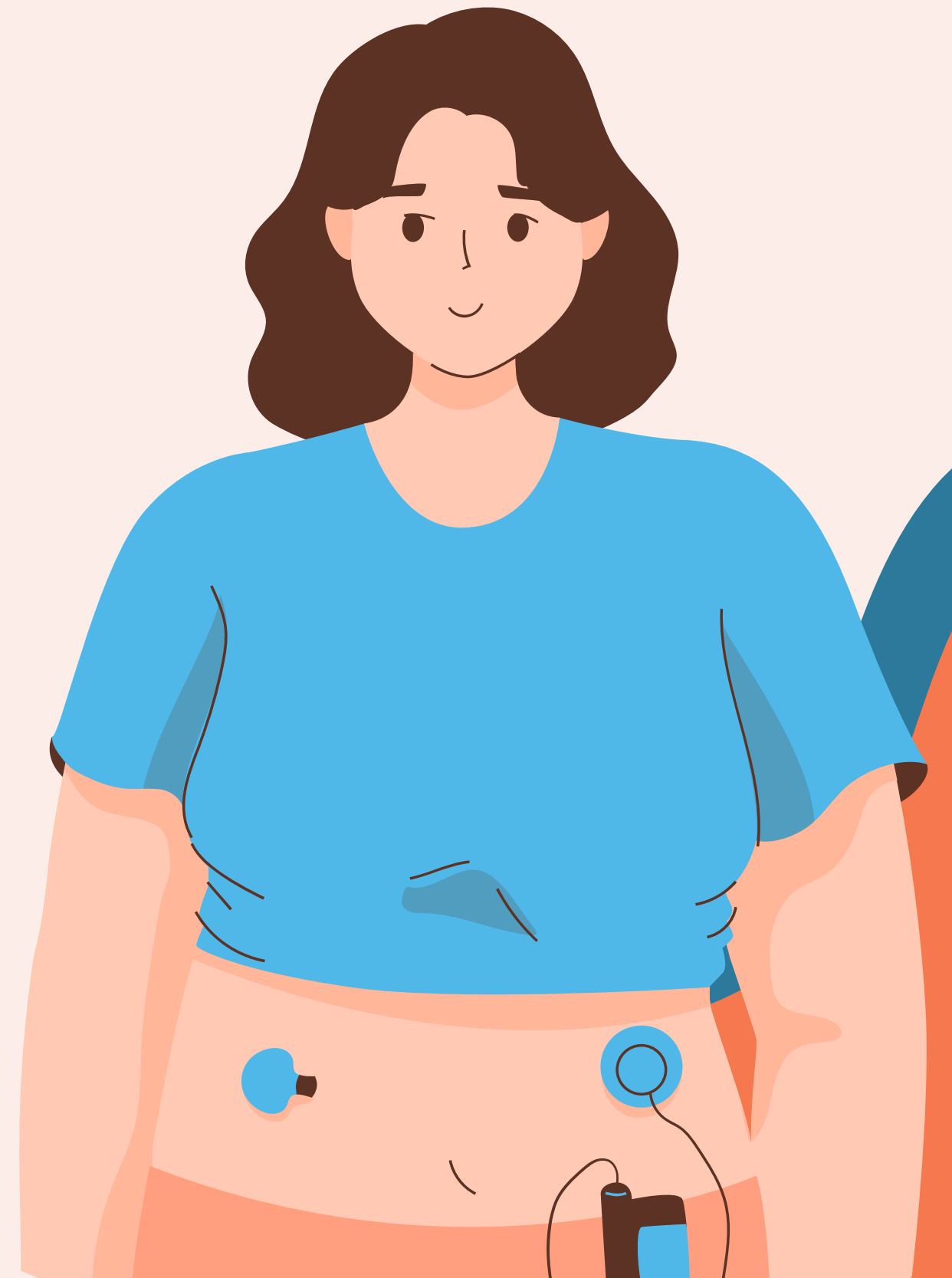


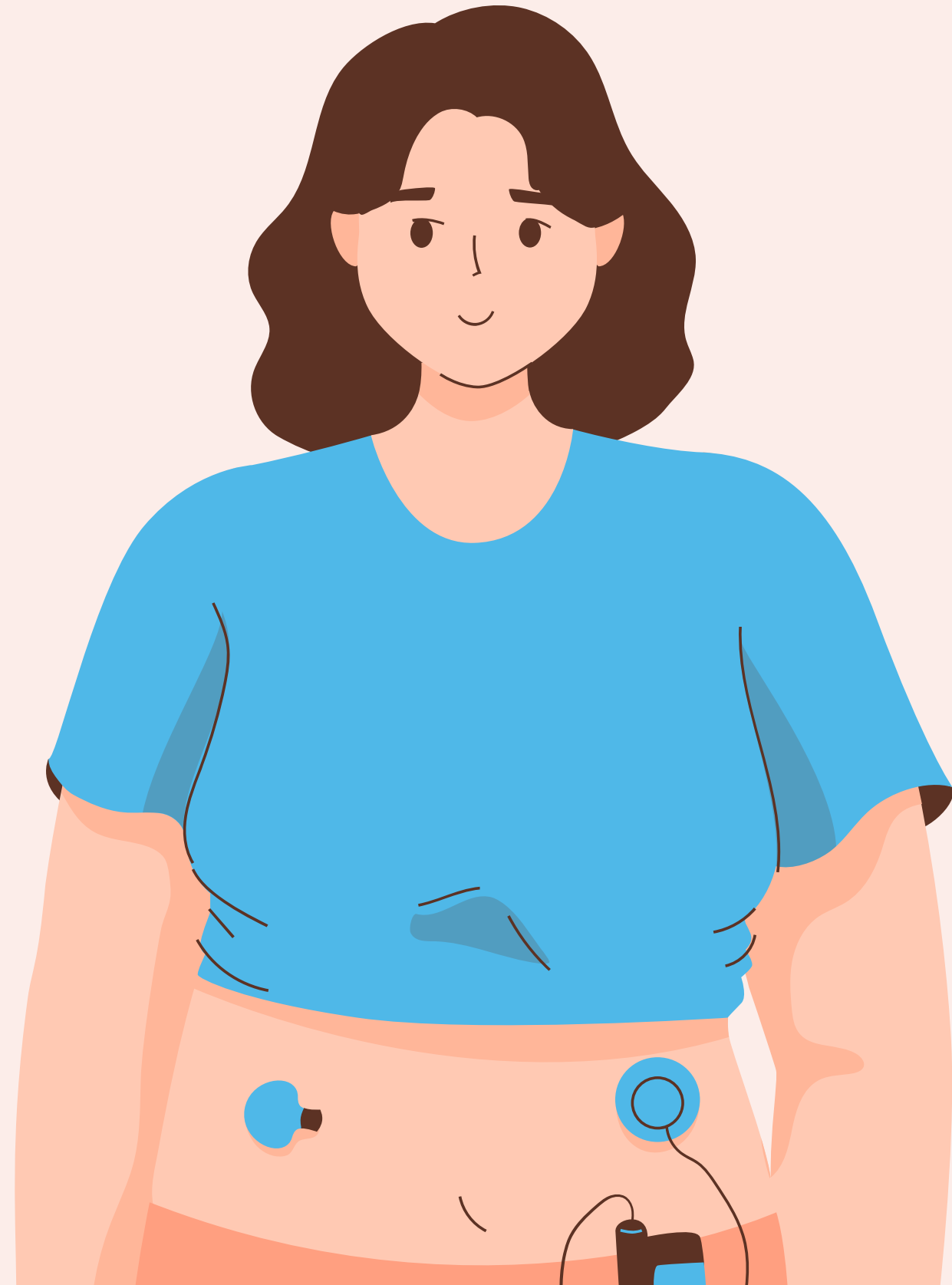
SARAN BISNIS

- Gunakan model KNN berbasis hasil KMeans untuk menyaring pasien berisiko diabetes secara otomatis.
- Dengan hasil KMeans clustering yang cukup akurat dalam mengelompokkan pasien, rumah sakit dapat:
 1. Membagi pasien ke dalam cluster risiko rendah dan tinggi.
 2. Menyusun program edukasi, pemantauan, dan perawatan berbeda berdasarkan segmen tersebut.
- Hasil clustering menunjukkan bahwa sebagian besar penderita diabetes bisa dikenali tanpa label eksplisit, artinya data historis bisa digunakan untuk skrining populasi.



THANK YOU!





WHAT ARE THE DIFFERENT TYPES OF DIABETES?



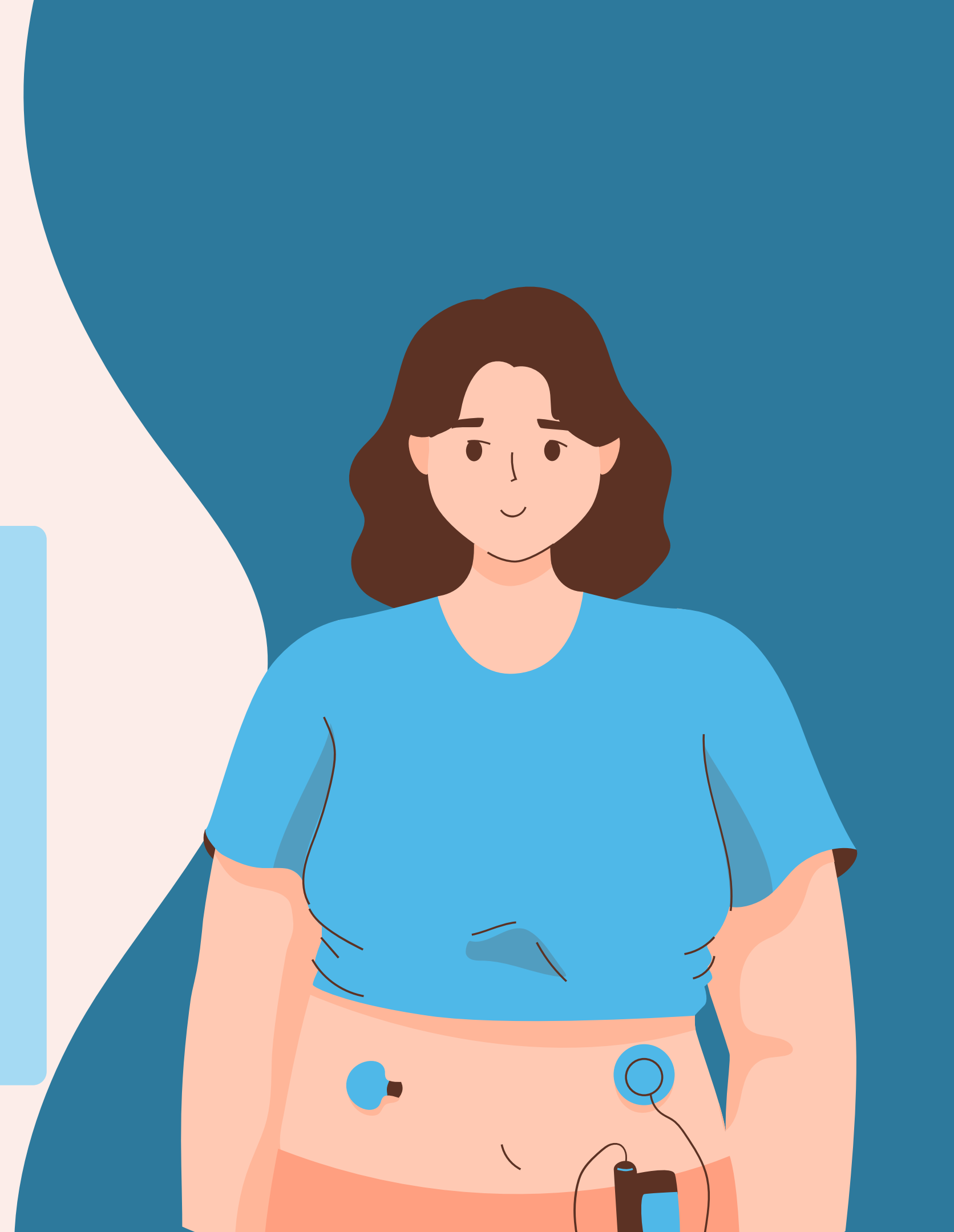
WHAT ABOUT SPORTS AND EXERCISE?

Explain how exercise can help manage diabetes.



WHAT TYPES OF MEDICATION WILL YOU NEED?

Explain the type of medications people with diabetes should take.



GET IN TOUCH

WEBSITE:

reallygreatsite.com

EMAIL ADDRESS:

hello@reallygreatsite.com

SOCIAL MEDIA:

[@reallygreatsite](#)

CONTACT NUMBER:

+123-456-7890