

DIABETES DISEASE PREDICTION

Tim Engineering membangun sistem cerdas yang mampu mendeteksi potensi penyakit diabetes pada pasien wanita di Amerika Utara (Pima Indians).



Diabetes Disease Prediction

NLP C – Alan turing

TABLE OF CONTENTS

- Background
- Business and Data Understanding
- Data Cleaning and Analysis
- Data Modeling
- Model Evaluation
- Results
- Conclusion



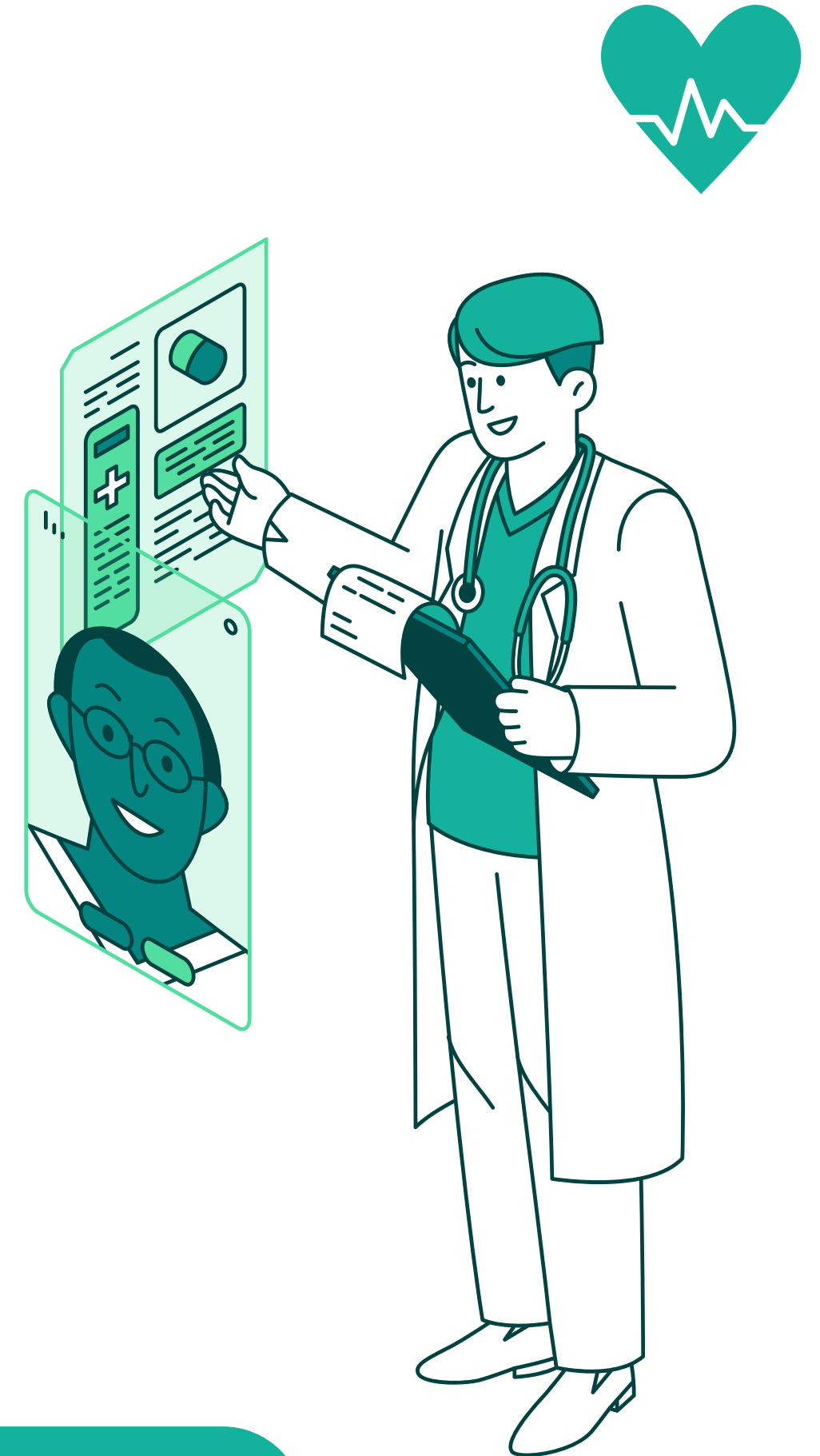
BACKGROUND

Diabetes adalah kondisi di mana tubuh tidak dapat memproduksi atau menggunakan insulin dengan efektif, menyebabkan tingginya kadar gula (glukosa) dalam darah. Hal ini terjadi ketika tubuh tidak mampu memproduksi atau menggunakan insulin dengan efektif. Deep Learning dipercaya dapat berperan penting untuk menangani diabetes yang berpotensi menyebabkan masalah kesehatan yang serius pada pasien jika tidak dikelola dengan baik.



GOALS

Memprediksi secara diagnostik apakah seorang pasien memiliki diabetes atau tidak



SEPERTI APA INDUSTRI HEALTH CARE DI INDONESIA?

Industri perawatan kesehatan di Indonesia menghadapi tantangan yang signifikan terkait penyakit tidak menular (NCD), termasuk diabetes.

Tantangan Diabetes di Indonesia:

1. Peningkatan NCD

Pada tahun 1990, NCD menyumbang 43 persen beban penyakit di Indonesia, dibandingkan dengan 49 persen untuk penyakit menular seperti tuberkulosis. Saat ini, NCD menyumbang 69 persen beban penyakit, dan angka ini terus meningkat.

- Cekindo. "Digitization in Indonesia Healthcare System: All You Need To Know"
- Mordor Intelligence. "Indonesia Diabetes Care Devices Market Size & Share Analysis"



SEPERTI APA INDUSTRI HEALTH CARE DI INDONESIA?

Tantangan Diabetes di Indonesia:

2. Biaya Besar

Tiga kategori NCD—penyakit kardiovaskular, diabetes beserta komplikasinya, dan penyakit pernapasan—menyebabkan hampir 50 persen kematian di negara ini. Diperkirakan Indonesia akan mengeluarkan sekitar 2,8 triliun dolar dari tahun 2012 hingga 2030 untuk mengatasi ketiga NCD ini, yang setara dengan tiga kali PDB Indonesia pada tahun 2014

- Cekindo. "Digitization in Indonesia Healthcare System: All You Need To Know"
- Mordor Intelligence. "Indonesia Diabetes Care Devices Market Size & Share Analysis"

Diabetes Disease Prediction

NLP C – Alan turing



SEBERAPA PENTING SEBUAH SISTEM CERDAS INI DIBANDINGKAN DENGAN PENDEKATAN DIAGNOSTIK YANG SUDAH ADA?

Metode yang sudah ada

Tes Gula Darah

Metode ini telah digunakan secara luas untuk mendiagnosis diabetes. Namun, fokus hanya pada diagnosis tanpa memperhatikan pencegahan dan manajemen jangka panjang dan dalam beberapa daerah terpencil belum mencakup metode ini.

- Jurnal Teknologi Terpadu: Klasifikasi penderita diabetes menggunakan algoritma machine learning
- BioMed Central: Artikel tentang machine learning dan deep learning algorithms untuk diabetes

Diabetes Disease Prediction

NLP C – Alan turing



SEBERAPA PENTING SEBUAH SISTEM CERDAS INI DIBANDINGKAN DENGAN PENDEKATAN DIAGNOSTIK YANG SUDAH ADA?

Solusi prediksi menggunakan machine learning memiliki signifikansi penting dalam mengatasi tantangan diabetes di Indonesia.

Dengan fokus pada pencegahan, efisiensi biaya, dan akses yang lebih baik, metode ini dapat membantu mengurangi risiko diabetes dan meningkatkan manajemen penyakit secara efektif.

Semua pendekatan, baik inovatif maupun yang sudah ada, harus diperkuat dengan kerjasama antara pemerintah, lembaga kesehatan, dan masyarakat untuk mengatasi beban diabetes secara efektif.

- Jurnal Teknologi Terpadu: Klasifikasi penderita diabetes menggunakan algoritma machine learning
- BioMed Central: Artikel tentang machine learning dan deep learning algorithms untuk diabetes



SEPERTI APA ARSITEKTUR NEURAL NETWORK YANG SESUAI UNTUK BISA MEMBANTU MENYELESAIKAN MASALAH YANG ADA?

Deep 1D-Convolutional Neural Network (DCNN)

Algoritma DCNN yang diusulkan berhasil dan unggul dalam mengatasi masalah diabetes melalui eksperimen pada dataset diabetes Pima Indian

<https://link.springer.com/article/10.1007/s00521-021-06431-7>

Sistem Diagnosis Diabetes dengan Artificial Neural Network

Arsitektur ini dapat membantu dalam mendiagnosis DR

<https://ieeexplore.ieee.org/document/9396963>

Model Prediksi Diabetes dengan Deep Neural Network (DNN):

DNN dapat membantu dalam mendiagnosis diabetes.

<https://jwcn-urasipjournals.springeropen.com/articles/10.1186/s13638-020-01765-7>

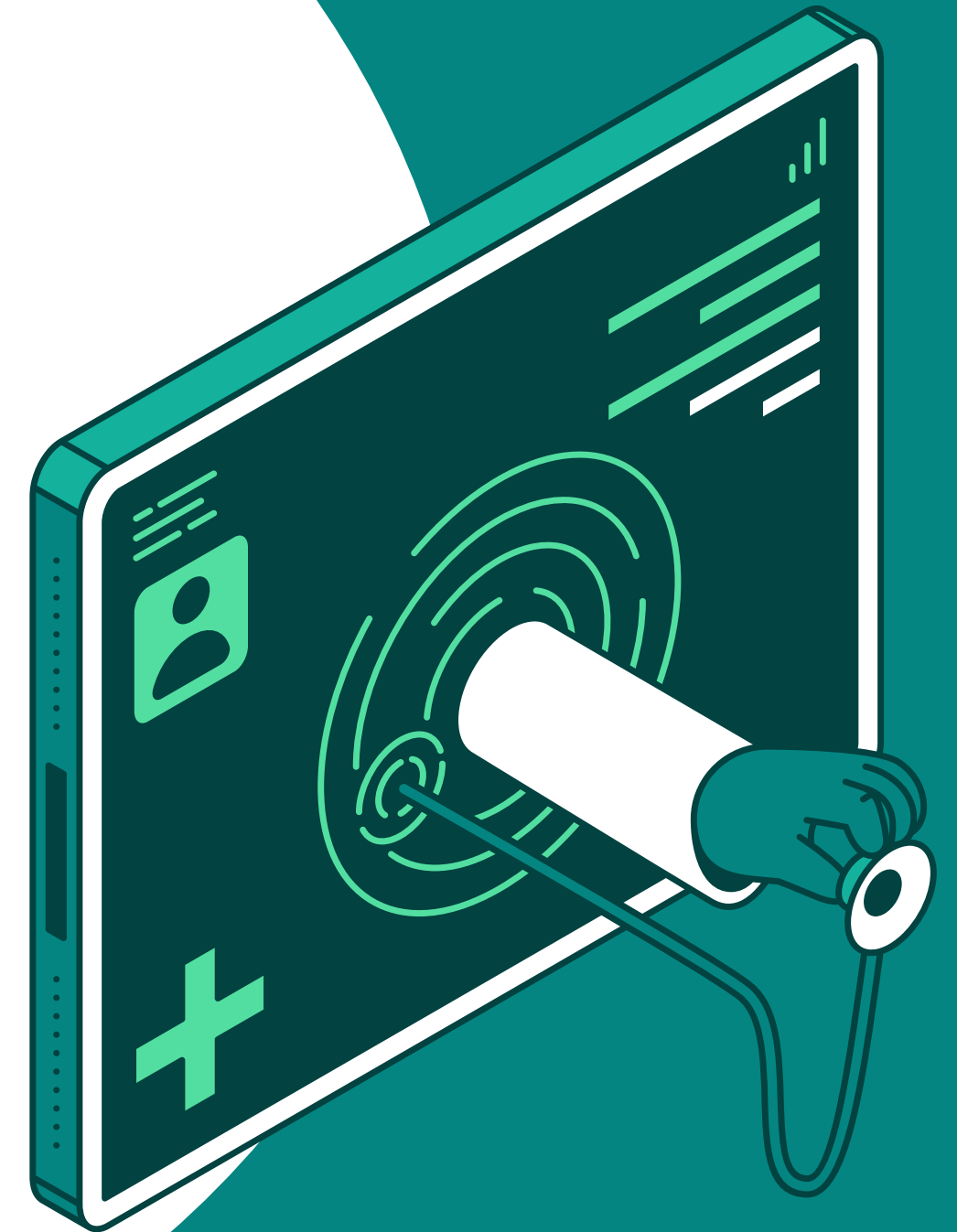


DATASET UNDERSTANDING

Pima Indians Diabetes Database

Kaggle

- Dataset terdiri atas 768 baris dan 9 kolom
- Dataset merupakan pasien perempuan, setidaknya berusia 21 tahun, berketurunan Indian Pima.
- Dataset berisi informasi medis pasien



DATASET UNDERSTANDING

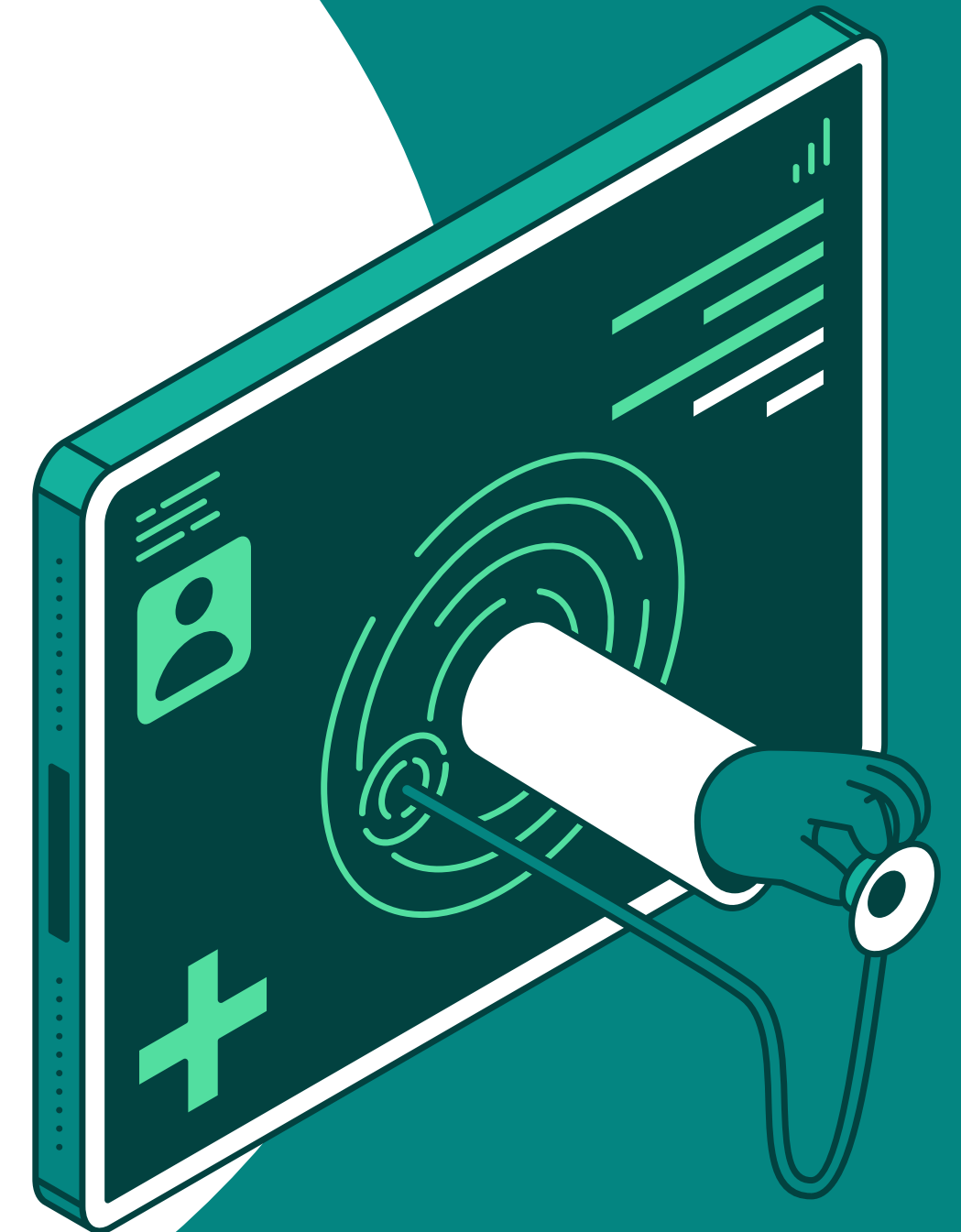
Fitur	Deskripsi
Pregnancies	Jumlah kehamilan yang pernah dialami oleh pasien.
Glucose	Kadar glukosa dalam plasma 2 jam setelah tes toleransi glukosa oral.
BloodPressure	Tekanan darah diastolik (tekanan saat jantung beristirahat di antara detak).
SkinThickness	Ketebalan lipatan kulit trisep.
Insulin	Kadar insulin serum 2 jam setelah tes toleransi glukosa oral.



DATASET UNDERSTANDING



Fitur	Deskripsi
BMI (Body Mass Index)	Indeks massa tubuh, pengukuran proporsi berat badan dan tinggi badan.
DiabetesPedigreeFunction	Skor fungsi keturunan diabetes, menggambarkan seberapa mungkin seseorang memiliki riwayat keturunan diabetes berdasarkan riwayat keluarga.
Age	Usia pasien.
Outcome	Kondisi keberadaan diabetes, di mana 1 menunjukkan keberadaan diabetes dan 0 menunjukkan ketiadaan diabetes.



MISSING VALUE



Terdapat 9 kolom dan 768 baris dimana 768 baris adalah non-null sehingga dapat disimpulkan tidak ada missing value

```
1 data = pd.read_csv('/work/diabetes.csv')
2 display(data.info(), data.head())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 768 entries, 0 to 767
Data columns (total 9 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Pregnancies            768 non-null   int64
1   Glucose                768 non-null   int64
2   BloodPressure          768 non-null   int64
3   SkinThickness          768 non-null   int64
4   Insulin                768 non-null   int64
5   BMI                   768 non-null   float64
6   DiabetesPedigreeFunction 768 non-null   float64
7   Age                   768 non-null   int64
8   Outcome                768 non-null   int64
dtypes: float64(2), int64(7)
memory usage: 54.1 KB
```

ANOMALY DATA



Beberapa fitur memiliki nilai 0 dan ini seharusnya tidak wajar. Maka dari itu, data anomali akan diimputasi

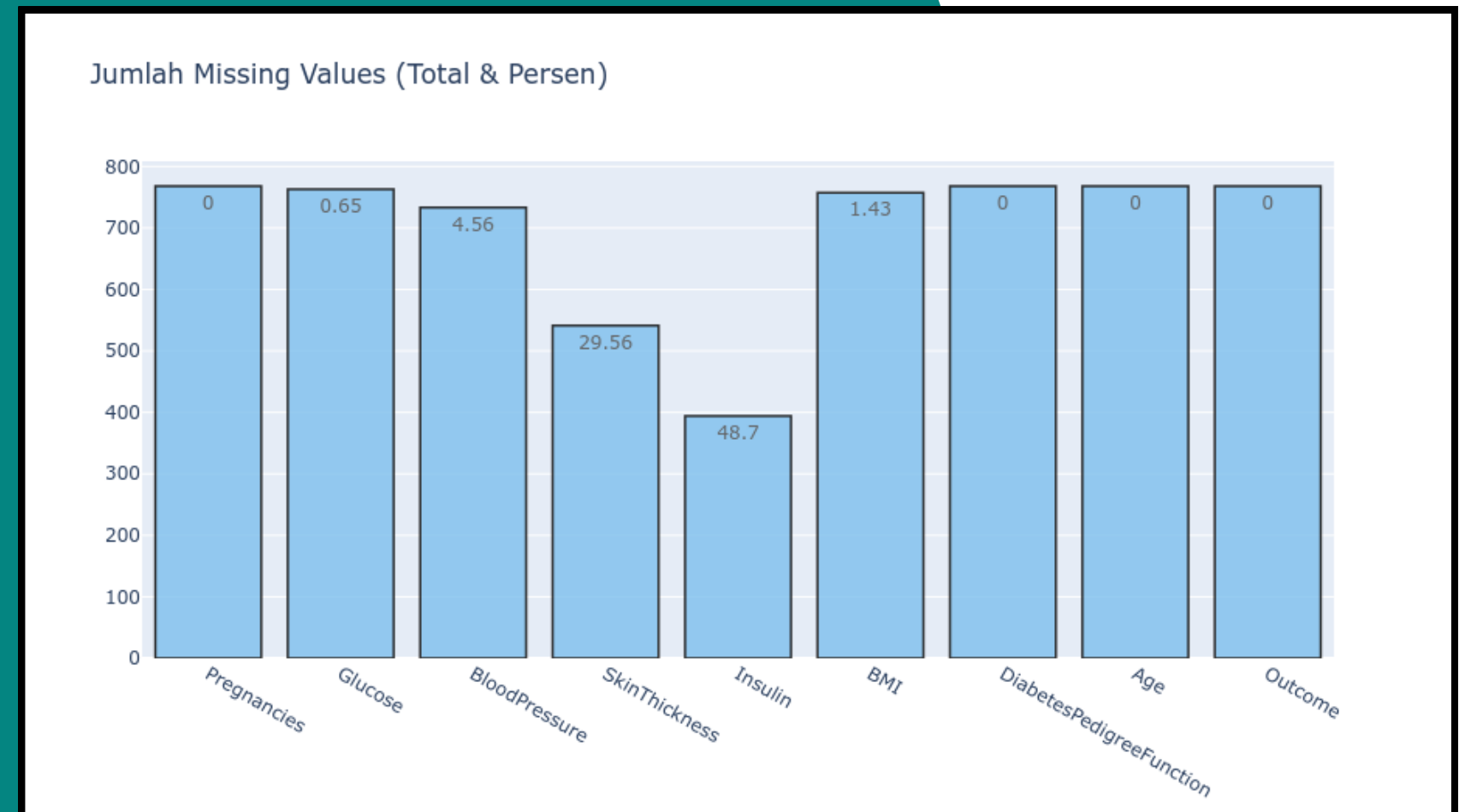
Insulin = 48.7% - 374

SkinThickness = 29.56% - 227

BloodPressure = 4.56% - 35

BMI = 1.43% - 11

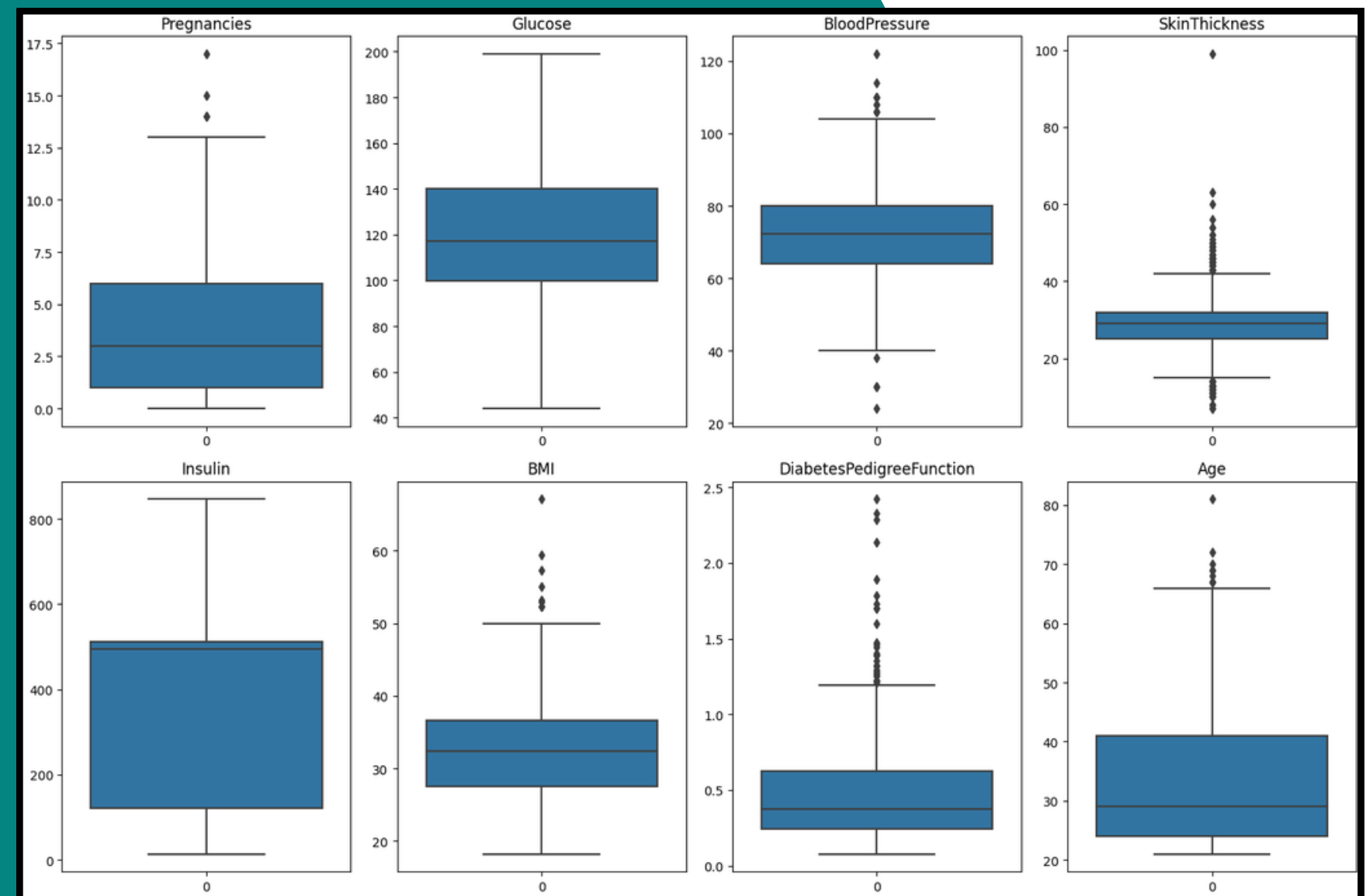
Glucose = 0.65% - 5



OUTLIER



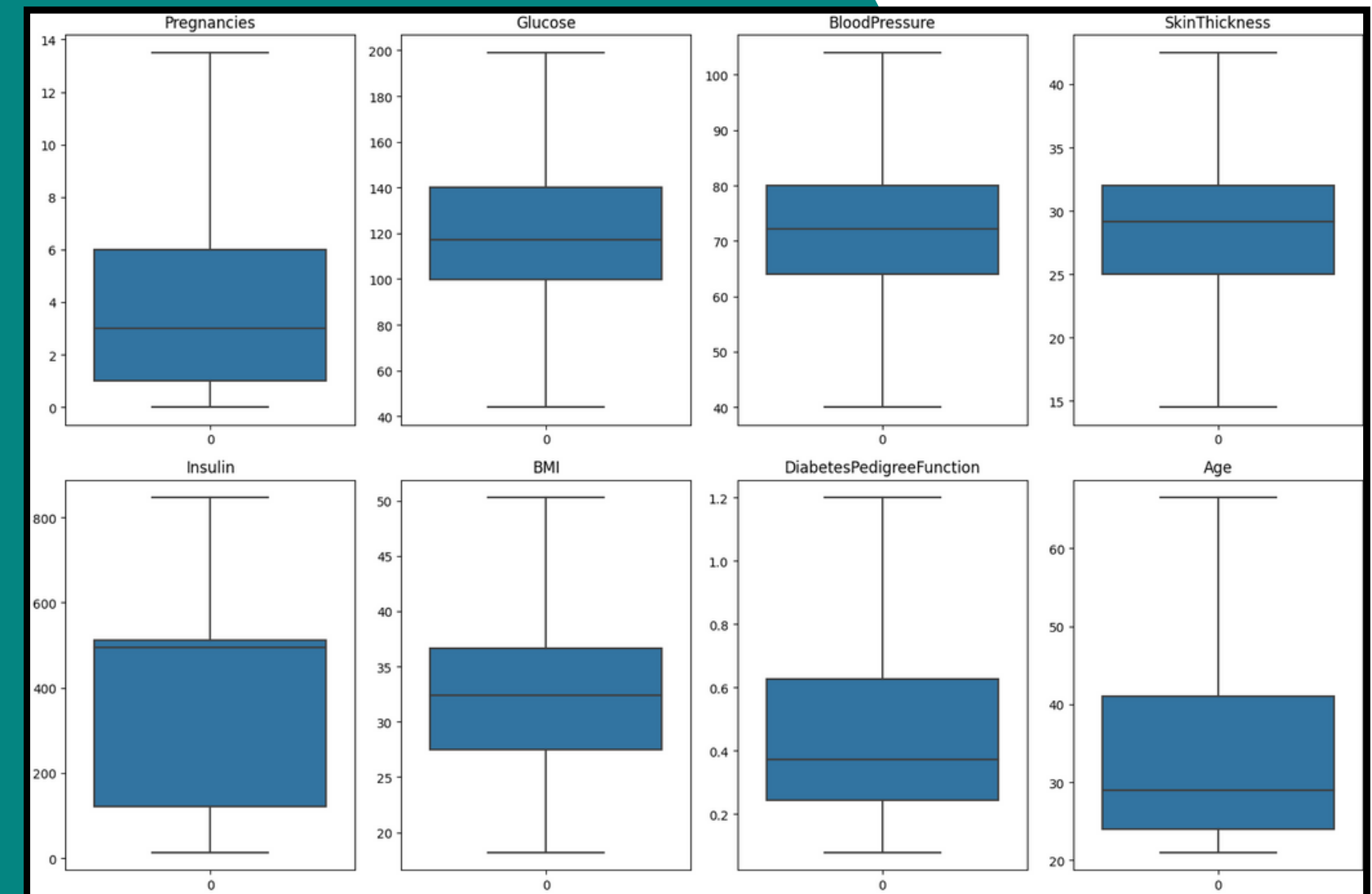
Terdapat beberapa fitur yang memiliki outlier tinggi yaitu Kolom DiabetesPedigreeFunction dan SkinThickness



OUTLIER HANDLING



Fitur dengan outlier yang tinggi akan di handle menggunakan IQR (Interquartile Range). Setelah di handle, dapat dilihat outlier sudah tidak terlihat pada BoxPlot



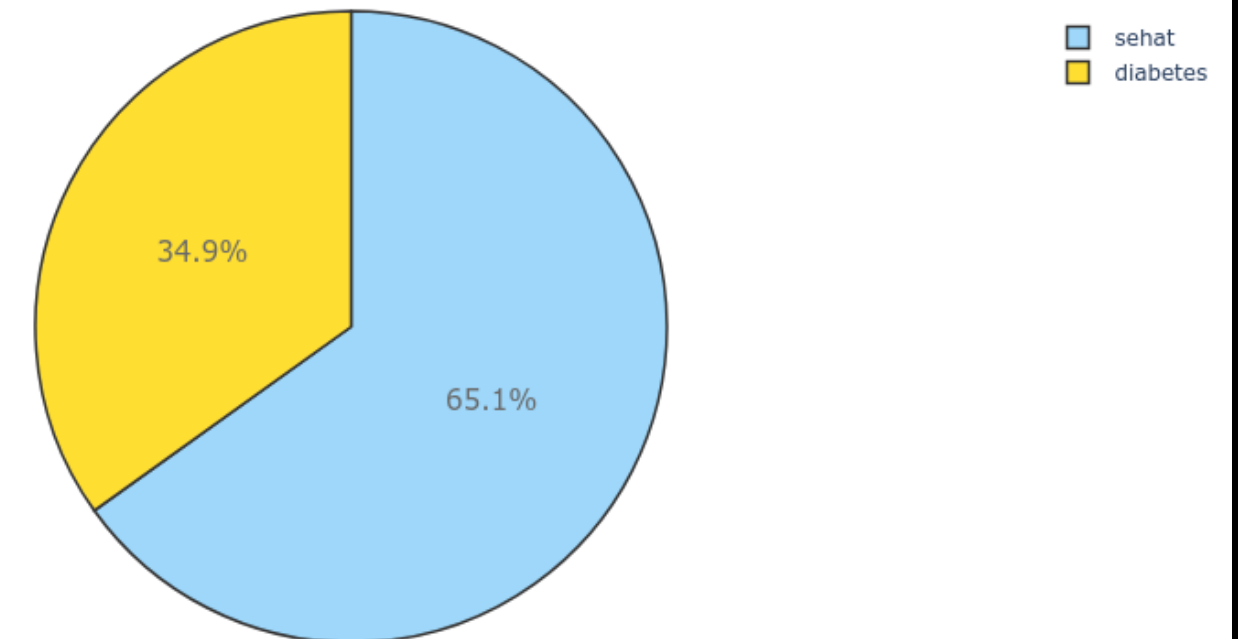
UNBALANCED DATA



Terdapat unbalance data target dengan Jumlah orang non-diabetes adalah 268 dan Jumlah pasien diabetes adalah 500.

Ketidakseimbangan ini akan diatasi dengan oversampling menggunakan Smote.

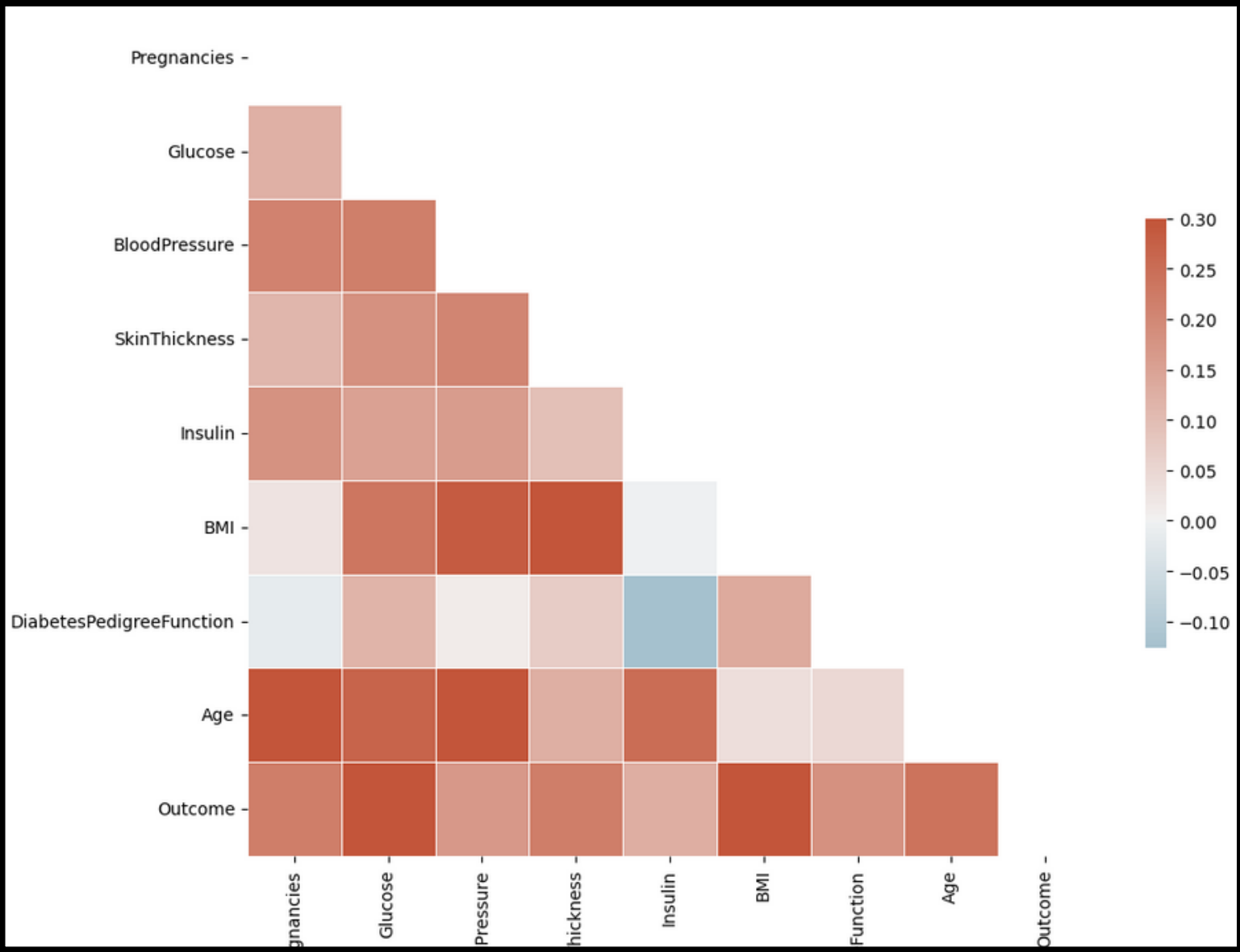
Distribusi



FEATURE CORRELATION



Fitur	Correlation
Pregnancies	0.22
Glucose	0.49
BloodPressure	0.17
SkinThickness	0.22
Insulin	0.13

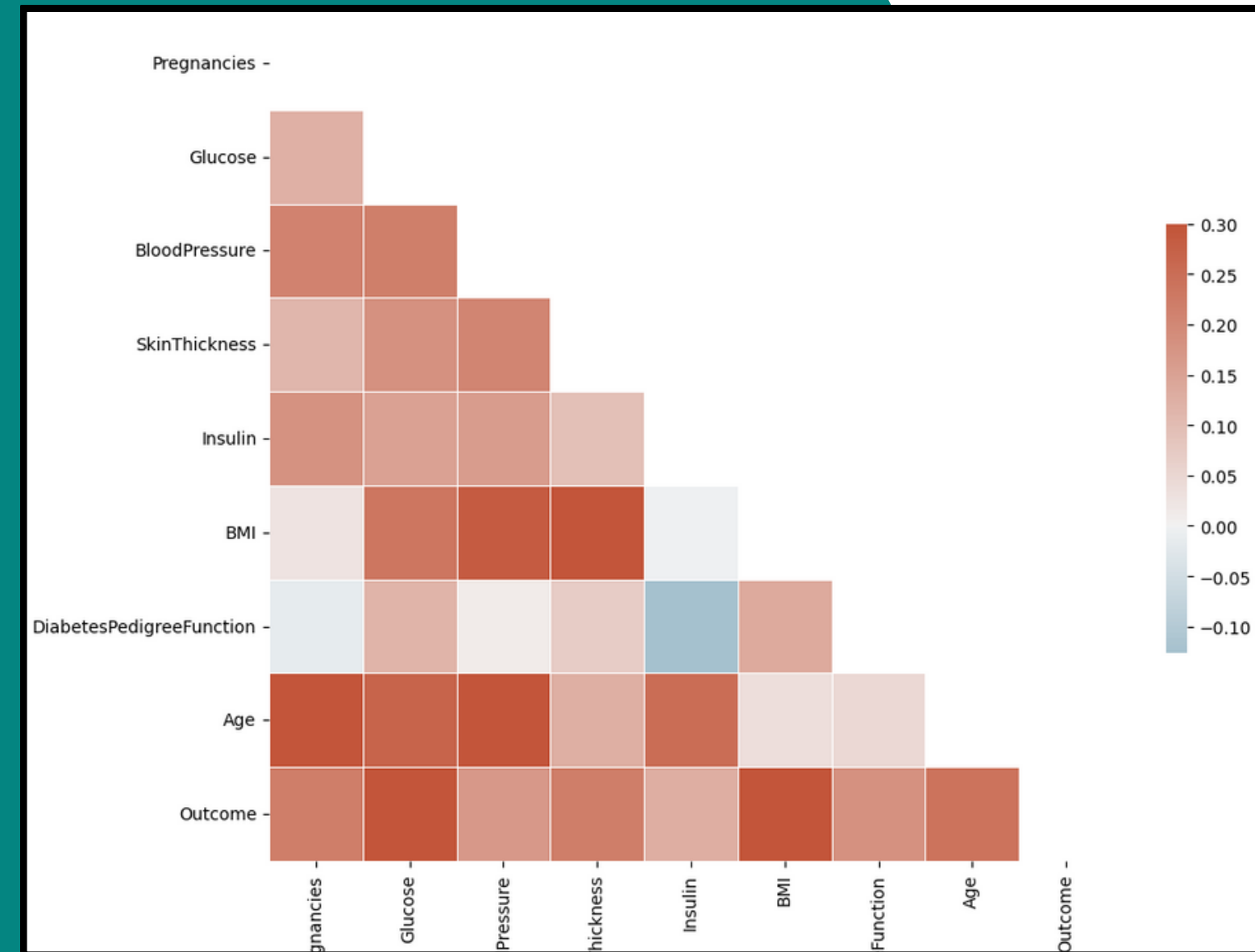


FEATURE CORRELATION



Fitur	Correlation
BMI	0.31
DiabetesPedigreeFunction	0.18
Age	0.24

Dataset memiliki fitur yang sedikit dan Setiap fitur memiliki korelasi yang cukup signifikan dengan target, maka dari itu semua fitur akan digunakan.



DATA MODELING



```
loss='binary_crossentropy', optimizer='adam', metrics=['accuracy']  
model_ds.fit(X_train_res, y_train_res, batch_size=64, epochs=1000, validation_split=0.3, callbacks=[checkpoint])
```

NARROW SHALLOW

```
model_ns = Sequential()  
model_ns.add(Dense(8, input_dim=8, activation='relu')) # Input layer  
model_ns.add(Dense(1, activation='sigmoid')) # Output layer
```

NARROW WIDE

```
model_nw = Sequential()  
model_nw.add(Dense(8, input_dim=8, activation='relu')) # Input layer  
model_nw.add(Dense(4, activation='relu'))  
model_nw.add(Dense(2, activation='relu'))  
model_nw.add(Dense(1, activation='sigmoid')) # Output layer
```

DEEP SHALLOW

```
model_ds = Sequential()  
model_ds.add(Dense(64, input_dim=8, activation='relu')) # Input layer  
model_ds.add(Dense(1, activation='sigmoid')) # Output layer
```

DEEP WIDE

```
model_dw = Sequential()  
model_dw.add(Dense(64, input_dim=8, activation='relu')) # Input layer  
model_dw.add(Dense(64, activation='relu'))  
model_dw.add(Dense(32, activation='relu'))  
model_dw.add(Dense(32, activation='relu'))  
model_dw.add(Dense(16, activation='relu'))  
model_dw.add(Dense(16, activation='relu'))  
model_dw.add(Dense(8, activation='relu'))  
model_dw.add(Dense(4, activation='relu'))  
model_dw.add(Dense(1, activation='sigmoid')) # Output layer
```

RESULT



- Narrow Shallow Test Accuracy
70.13 %
- Narrow Wide Test Accuracy
67.53 %



- Deep Shallow Test Accuracy :
68.83 %
- Deep Wide Test Accuracy :
68.83 %



CONCLUSION

Dalam proyek prediksi diabetes, model Neural Network (ANN) berhasil mencapai akurasi pada data test 70.13% menggunakan arsitektur narrow shallow yang sederhana. Hasil ini dapat memenuhi goals diagnostik untuk memprediksi apakah seorang pasien memiliki diabetes atau tidak. Dengan akurasi yang memadai, model memberikan informasi berharga dalam pengambilan keputusan klinis terkait diagnosis diabetes pada pasien.