Prediksi Nilai Systolic dan Diastolic berdasarkan Sinyal PPG

A diagram of a bar chart

Description automatically generated with medium confidence

Zufar Asyraf Al Hamid   
201012310075

# Hasil

Dalam percobaan ini, dilakukan Exploratory Data Analysis (EDA) untuk mendapatkan pemahaman awal tentang distribusi dan karakteristik dari. Proses EDA meliputi:

* Visualisasi Data: Membuat Histogram untuk melihat distribusi Mean\_NBP\_Sys dan Mean\_NBP\_Dias, yang menunjukkan sebaran data dan membantu dalam identifikasi outlier seperti pada gambar 1.
* A diagram of a number of blue dots

  Description automatically generatedStatistik Deskriptif: Menampilkan statistik dasar seperti mean, standar deviasi, minimum, dan maksimum yang memberikan insight tentang pusat dan penyebaran data seperti pada gambar 2.
* Analisis Boxplot: Menampilkan boxplot untuk melihat distribusi data lebih lanjut dan mengidentifikasi outlier terlihat pada gambar 3.
* A graph of a number of blue and white bars

  Description automatically generated with medium confidenceScatter Plot: Membuat plot hubungan antara Mean\_NBP\_Sys dan Mean\_NBP\_Dias untuk menilai korelasi antara kedua variabel target terlihat pada gambar 4.

A screenshot of a computer screen

Description automatically generated

Setelah melakukan EDA, diterapkannya proses train dan testing model. Di sini digunakan tiga buah model untuk dapat membandingkan performasi antar model. Model yang digunakan adalah Regresi Linear, Random Forest, dan KNN. Sebelum memasuki pelatihan, data dinormalisasi menggunakan standar scaler agar dapat dilakukan pelatihan. Setelah itu data dibagi menjadi 80% training dan 20% testing. Hasil dari ketiga model adalah sebgai berikut:

* Linear Regression:

RMSE: 36.9599

R2: -11.5149

* Random Forest:

RMSE: 11.0188

R2: -0.1157

* K-Nearest Neighbors (KNN):

RMSE: 11.6206

R2: -0.2217

Dari ketiga model tersebut, Random Forest menunjukkan RMSE yang lebih rendah sebelum proses tuning, menandakan performa yang lebih baik dalam memprediksi tekanan darah. Hyperparameter tuning dilakukan untuk model Random Forest dengan menggunakan teknik Grid Search CV, yang menyertakan 144 kombinasi parameter dan dilakukan sebanyak 720 iterasi fitting. Hasilnya, kombinasi parameter terbaik adalah max\_depth: None, min\_samples\_leaf: 4, min\_samples\_split: 2, dan n\_estimators: 200. Setelah dilakukan tuning, Random Forest menunjukkan peningkatan dengan RMSE sebesar 10.6128 dan R2 sebesar -0.0474, meskipun R2 masih menunjukkan bahwa model tidak mampu menjelaskan variabilitas dari data dengan baik.

# Diskusi

Diskusi dalam laporan ini berfokus pada interpretasi dari hasil yang didapatkan. Kinerja awal Linear Regression yang rendah dapat dikaitkan dengan asumsi dasar regresi linier yang mungkin tidak sesuai dengan kompleksitas data. Random Forest yang menghasilkan hasil yang lebih baik menunjukkan bahwa metode berbasis *decision tree* lebih cocok untuk menangani variabilitas dalam data.

Namun, R2 yang negatif pada awalnya menunjukkan bahwa model-model ini belum mampu memprediksi dengan akurat, yang bisa disebabkan oleh fitur-fitur yang tidak cukup informatif atau kebutuhan akan pendekatan pemodelan yang lebih kompleks. Hasil hyperparameter tuning yang menunjukkan perbaikan pada nilai RMSE mengindikasikan bahwa optimasi parameter model adalah langkah penting yang bisa meningkatkan performa prediksi.

Kesimpulan dari diskusi ini adalah bahwa terdapat potensi untuk meningkatkan model dengan eksplorasi feature engineering lebih lanjut, pemilihan model yang lebih canggih, atau menggunakan ensemble learning untuk mengkombinasikan kekuatan dari berbagai model.