

LAPORAN MACHINE LEARNING



Anggota Kelompok:

- | | |
|------------------------|----------|
| 1. Made Yoga Mahardika | 26416083 |
| 2. Jong, Jeffrey W | 26416085 |
| 3. Nicolas Wiyendi | 26416088 |
| 4. Ricky Setiawan | 26416095 |

Data Pre-Processing

Proses tambahan yang dilakukan berupa:

- Mengubah data menjadi berskala 0 hingga 1. Caranya yaitu dengan menggunakan library MinMaxScaler.

Metode/Learning Algorithm

Metode yang digunakan ada 4, yaitu:

- K Nearest
- Naive Bayes Gaussian
- Decision Tree
- Neural Network (Multi Layer Precrepton)

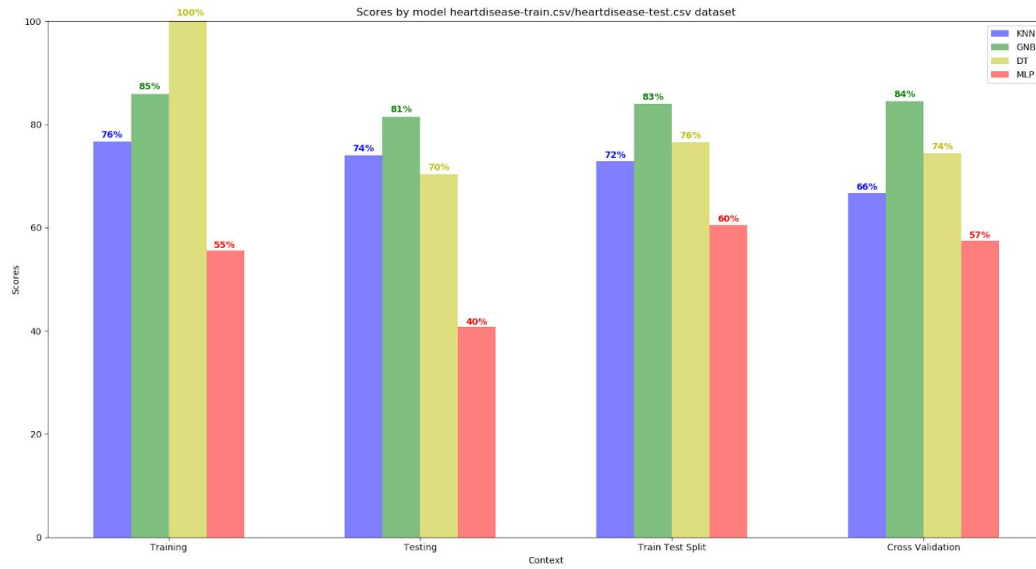
Accuracy

Hasil klasifikasi dari dataset:

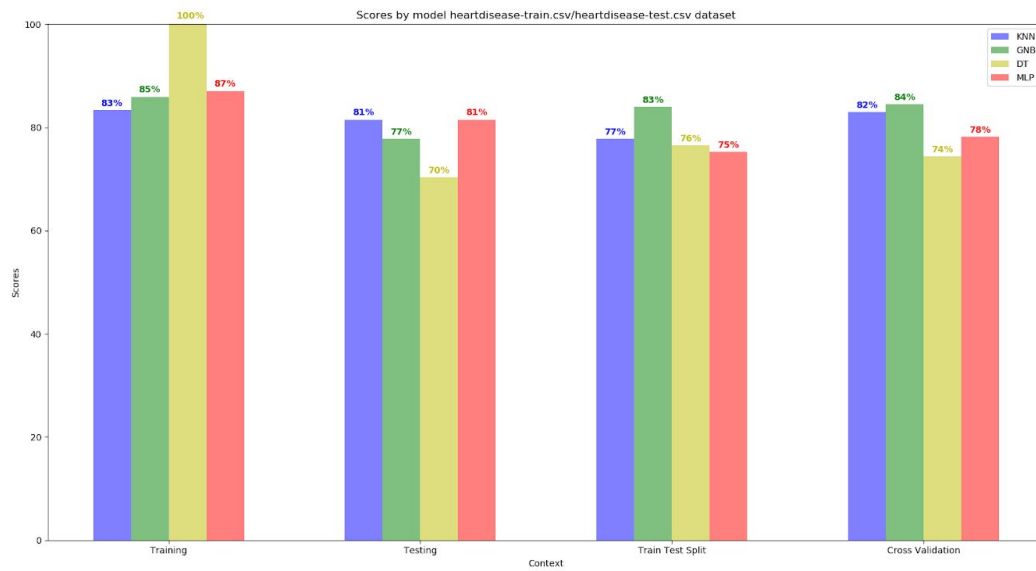
- **Training Accuracy:**
 - **Not Scaled:**
 - **Heart Disease:**
 - K-Nearest : 76%
 - Naives Bayes Gaussian : 85%
 - Decision Tree : 100%
 - Neural Network : 55%
 - **Iris:**
 - K-Nearest : 97%
 - Naives Bayes Gaussian : 96%
 - Decision Tree : 100%
 - Neural Network : 33%
 - **Scaled:**
 - **Heart Disease:**
 - K-Nearest : 83%
 - Naives Bayes Gaussian : 85%
 - Decision Tree : 100%
 - Neural Network : 87%
 - **Iris:**
 - K-Nearest : 96%
 - Naives Bayes Gaussian : 96%
 - Decision Tree : 100%
 - Neural Network : 100%
- **Test Accuracy:**
 - **Not Scaled:**
 - **Heart Disease:**
 - K-Nearest : 74%

- Naives Bayes Gaussian : 81%
 - Decision Tree : 70%
 - Neural Network : 40%
- **Iris:**
 - K-Nearest : 93%
 - Naives Bayes Gaussian : 93%
 - Decision Tree : 86%
 - Neural Network : 33%
- **Scaled:**
 - **Heart Disease:**
 - K-Nearest : 81%
 - Naives Bayes Gaussian : 77%
 - Decision Tree : 70%
 - Neural Network : 81%
 - **Iris:**
 - K-Nearest : 93%
 - Naives Bayes Gaussian : 93%
 - Decision Tree : 86%
 - Neural Network : 93%
- **Testing accuracy with Cross validation 10 cv:**
 - **Not Scaled:**
 - **Heart-Disease:**
 - K-Nearest : 66%
 - Naives Bayes Gaussian : 84%
 - Decision Tree : 74%
 - Neural Network : 57%
 - **Iris:**
 - K-Nearest : 95%
 - Naives Bayes Gaussian : 95%
 - Decision Tree : 95%
 - Neural Network : 33%
 - **Scaled:**
 - **Heart-Disease:**
 - K-Nearest : 82%
 - Naives Bayes Gaussian : 84%
 - Decision Tree : 74%
 - Neural Network : 78%
 - **Iris:**
 - K-Nearest : 97%
 - Naives Bayes Gaussian : 95%
 - Decision Tree : 95%
 - Neural Network : 96%

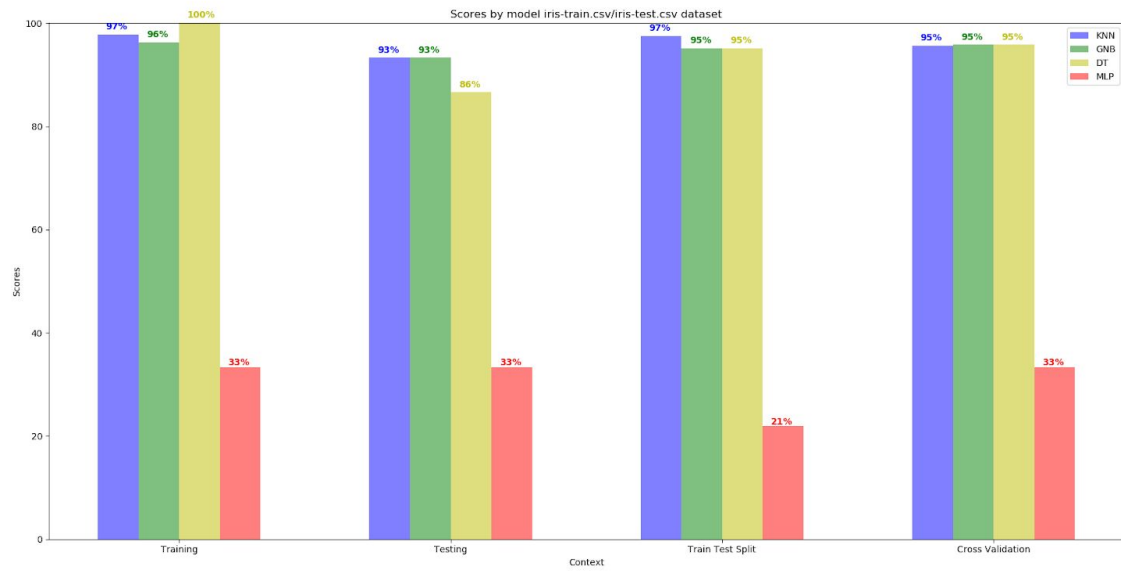
Heart Disease Accuracy Not Scaled



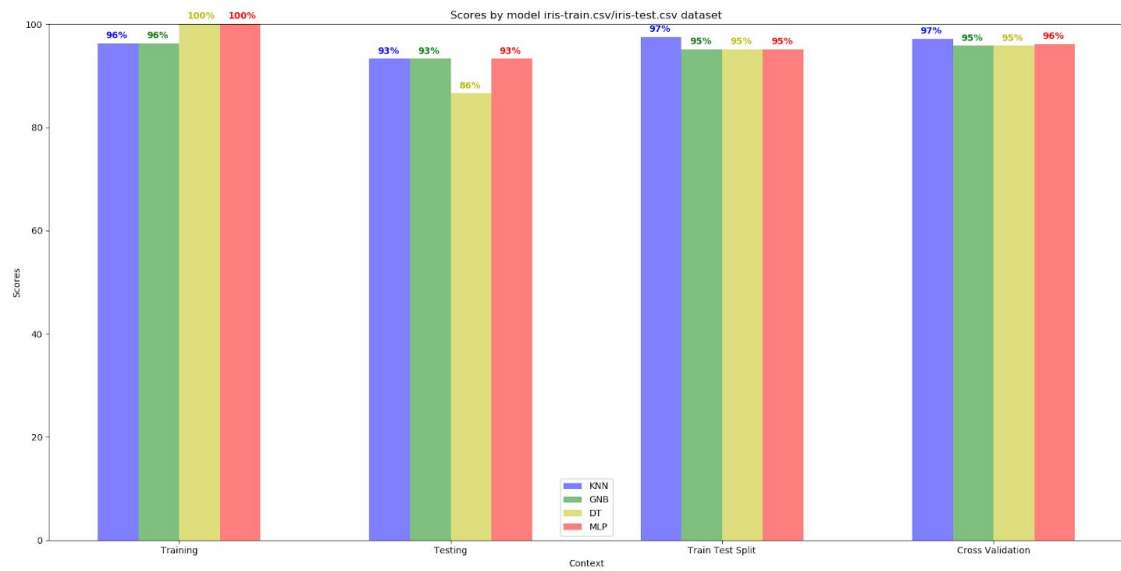
Heart Disease Accuracy Scaled



Iris Accuracy Not Scaled



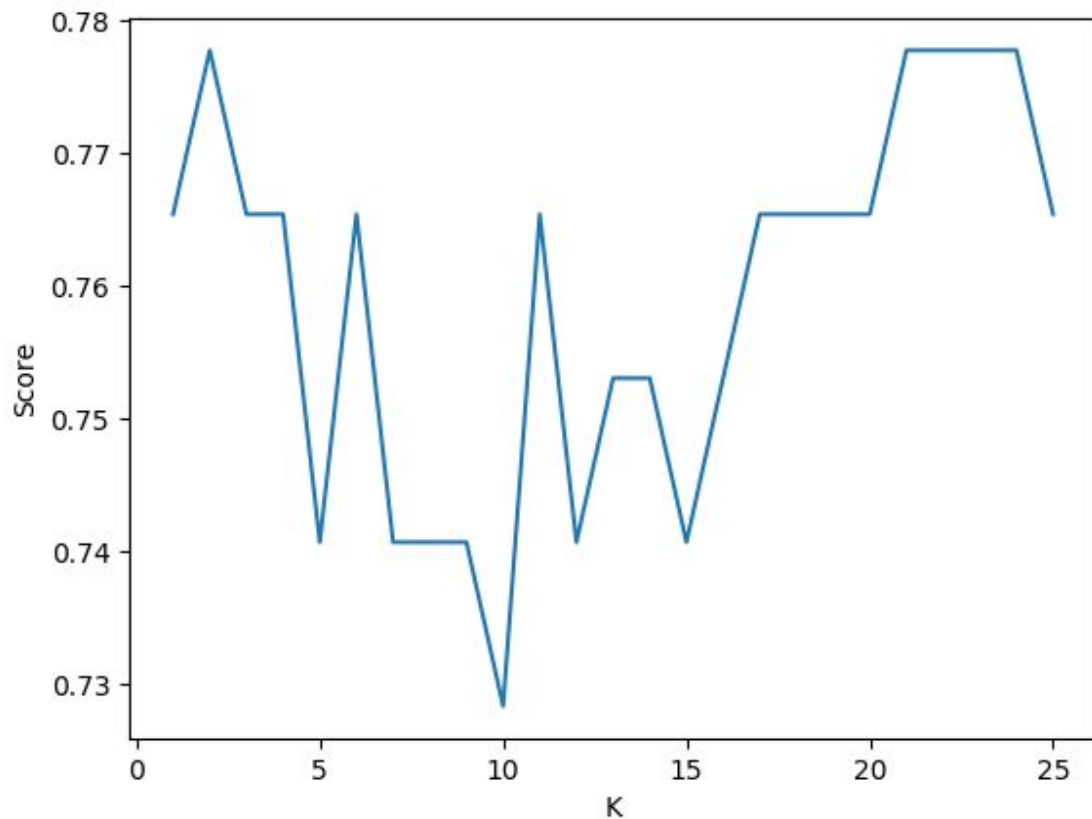
Iris Accuracy Scaled



Eksperimen

Eksperimen yang dilakukan dengan parameter dari algoritma yang digunakan:

- Neural network sangat sensitif terhadap data yang nilai parameternya tidak di *scale*.
- K Nearest Neighbors, jumlah K yang berbeda dari 1 - 25.



Keterangan lain

- **Bahasa pemrograman:**
 - Python
- **Library:**
 - Pandas:
 - Digunakan untuk membuka file berekstensi .csv.
 - Fitur selection.
 - Scikit-Learn:
 - Library utama dari metode yang digunakan.
 - Numpy:
 - Fitur selection.

Kesimpulan

- Untuk data Heart Disease lebih baik menggunakan model Naives Bayes Gaussian/Neural Network.
- Untuk data Iris lebih baik menggunakan model Naives Bayes Gaussian/Neural Network.
- Tetapi neural network, scaling data sangat berpengaruh/sensitif. Dapat dilihat di gambar grafik diatas. Accuracy akan berpengaruh pada data yang sudah discale.
- Diliat dari data graphic di atas, model Decision Tree mengalami Overfitting. Karena, Training Accuracy dan Test Accuracy dari model Decision Tree memiliki jarak nilai yang cukup besar.