

# Pembersihan Data

## 1. Oil-Spill Dataset

- a. Tanpa Pembersihan
  - i. Logistic Regression
    - Sensitivity = 0.9898648648648649
    - Specificity = 0.14285714285714285
    - Precision = 0.9606557377049181
    - NPV = 0.4
    - Accuracy = 0.9516129032258065
  - ii. MLP
    - Sensitivity = 0.9833333333333333
    - Specificity = 0.0
    - Precision = 0.9672131147540983
    - NPV = 0.0
    - Accuracy = 0.9516129032258065
- b. Penghapusan single value
  - i. Logistic Regression
    - Sensitivity= 0.9898648648648649
    - Specificity= 0.14285714285714285
    - Precision= 0.9606557377049181
    - NPV= 0.4
    - Accuracy= 0.9516129032258065
  - ii. MLP
    - Sensitivity= 0.9832214765100671
    - Specificity= 0.0
    - Precision= 0.9606557377049181
    - NPV= 0.0
    - Accuracy= 0.9451612903225807
- c. Penghapusan few values
  - i. Logistic Regression
    - Sensitivity= 0.9898648648648649
    - Specificity= 0.14285714285714285
    - Precision= 0.9606557377049181
    - NPV= 0.4
    - Accuracy= 0.9516129032258065
  - ii. MLP
    - Sensitivity= 0.9831081081081081
    - Specificity= 0.0
    - Precision= 0.9540983606557377
    - NPV= 0.0
    - Accuracy= 0.9387096774193548
- d. Penghapusan low variance
  - i. Logistic Regression

- Sensitivity= 0.9898648648648649
- Specificity= 0.14285714285714285
- Precision= 0.9606557377049181
- NPV= 0.4
- Accuracy= 0.9516129032258065
- ii. MLP
  - Sensitivity= 0.9833887043189369
  - Specificity= 0.0
  - Precision= 0.9704918032786886
  - NPV= 0.0
  - Accuracy= 0.9548387096774194
- e. Menggunakan Variance treshhold
  - \*Tercantum dalam file .ipnyb

## 2. Iris Dataset

- a. Tanpa Pembersihan
  - i. Logistic Regression
    - Iris-Setosa
      - a. Sensitivity = 1.0
      - b. Specificity = 1.0
      - c. Precision = 1.0
      - d. NPV = 1.0
      - e. Accuracy = 1.0
    - Iris-Versicolor
      - a. Sensitivity = 0.9047619047619048
      - b. Specificity = 0.9655172413793104
      - c. Precision = 0.95
      - d. NPV = 0.9333333333333333
      - e. Accuracy = 0.94
    - Iris-Virginica
      - a. Sensitivity = 0.9230769230769231
      - b. Specificity = 0.9459459459459459
      - c. Precision = 0.8571428571428571
      - d. NPV = 0.9722222222222222
      - e. Accuracy = 0.94
  - ii. MLP
    - Iris-Setosa
      - a. Sensitivity = 1.0
      - b. Specificity = 1.0
      - c. Precision = 1.0
      - d. NPV = 1.0
      - e. Accuracy = 1.0
    - Iris-Versicolor
      - a. Sensitivity = 0.625

- b. Specificity = 0.9705882352941176
    - c. Precision = 0.9090909090909091
    - d. NPV = 0.8461538461538461
    - e. Accuracy = 0.86
  - Iris-Virginica
    - a. Sensitivity = 0.9444444444444444
    - b. Specificity = 0.8125
    - c. Precision = 0.7391304347826086
    - d. NPV = 0.9629629629629629
    - e. Accuracy = 0.86
- b. Penghapusan Data Terduplikasi
  - i. Logistic Regression
    - Iris-Setosa
      - a. Sensitivity = 1.0
      - b. Specificity = 1.0
      - c. Precision = 1.0
      - d. NPV = 1.0
      - e. Accuracy = 1.0
    - Iris-Versicolor
      - a. Sensitivity = 0.625
      - b. Specificity = 0.9705882352941176
      - c. Precision = 0.9090909090909091
      - d. NPV = 0.8461538461538461
      - e. Accuracy = 0.86
    - Iris-Virginica
      - a. Sensitivity = 0.9444444444444444
      - b. Specificity = 0.8125
      - c. Precision = 0.7391304347826086
      - d. NPV = 0.9629629629629629
      - e. Accuracy = 0.86
  - ii. MLP
    - Iris-Setosa
      - a. Sensitivity = 1.0
      - b. Specificity = 1.0
      - c. Precision = 1.0
      - d. NPV = 1.0
      - e. Accuracy = 1.0
    - Iris-Versicolor
      - a. Sensitivity = 0.75
      - b. Specificity = 1.0
      - c. Precision = 1.0
      - d. NPV = 0.8529411764705882
      - e. Accuracy = 0.8979591836734694
    - Iris-Virginica

- a. Sensitivity = 1.0
- b. Specificity = 0.8529411764705882
- c. Precision = 0.75
- d. NPV = 1.0
- e. Accuracy = 0.8979591836734694

Analisa :

Dengan membersihkan data kita akan meningkatkan produktifitas dan kualitas informasi yang model kita prediksi. Dapat dilihat melalui data-data diatas bahwa secara rata-rata terdapat kenaikan nilai Sensitivity, Specificity, Precision, NPV, maupun Accuracy. Contohnya pada model MLP data Iris-Virginica, Accuracy yang sebelumnya 0.86 meningkat menjadi 0.8979591836734694.