**Mudah Belajar Otodidak Data Science**

**(Praktek menggunakan python3)**

**Disusun oleh**

**Aryajaya Alasmsyah, S.Kom., M.Kom., MTA.**

**Pertemuan 4 dan 5 – Algoritma Klasifikasi Data**

**Tujuan pembelajaran**

* Mahasiswa mampu memahami beberapa algoritma klasifikasi seperti decision tree, naïve bayes, k-nearest neighbor, support vector machine dan logistik regresi.
* Mahasiswa mampu memahami perbedaan antara algoritma klasifikasi decision tree, naïve bayes, k-nearest neighbor, support vector machine dan logistik regresi.

**Studi kasus: Klasifikasi Bunga Iris Menggunakan Algoritma ID3, C45, GNB, MNB, KNN, SVM, dan LR.**

|  |
| --- |
| # Lib untuk manipulasi data  import numpy as np  import pandas as pd    # lib untuk analisa statistik  import statsmodels.api as sm    # lib untuk visualisasi data  import seaborn as sns  import matplotlib.pyplot as plt  import plotly.express as px  import plotly.graph\_objects as go    # lib untuk praproses dataset  from sklearn.preprocessing import MinMaxScaler  from sklearn.model\_selection import train\_test\_split    # lib untuk klasifikasi data  from sklearn.tree import DecisionTreeClassifier  from sklearn.neighbors import KNeighborsClassifier  from sklearn.naive\_bayes import GaussianNB, MultinomialNB, BernoulliNB  from sklearn.linear\_model import LogisticRegression  from sklearn.svm import SVC    # library untuk evaluasi model  from sklearn.metrics import accuracy\_score, precision\_score, recall\_score, f1\_score  from sklearn.metrics import classification\_report, confusion\_matrix |

**1. Data Acquisition**

|  |
| --- |
| # load dataset  dataset = pd.read\_csv("../dataset/iris.csv")    # show metadata of dataset  print(dataset.info()) |
| <class 'pandas.core.frame.DataFrame'>  RangeIndex: 150 entries, 0 to 149  Data columns (total 5 columns):  # Column Non-Null Count Dtype  --- ------ -------------- -----  0 sepal\_length 150 non-null float64  1 sepal\_width 150 non-null float64  2 petal\_length 150 non-null float64  3 petal\_width 150 non-null float64  4 species 150 non-null object  dtypes: float64(4), object(1)  memory usage: 6.0+ KB |

|  |
| --- |
| # show dataset  print(dataset) |
| sepal\_length sepal\_width petal\_length petal\_width species  0 5.1 3.5 1.4 0.2 setosa  1 4.9 3.0 1.4 0.2 setosa  2 4.7 3.2 1.3 0.2 setosa  3 4.6 3.1 1.5 0.2 setosa  4 5.0 3.6 1.4 0.2 setosa  .. ... ... ... ... ...  145 6.7 3.0 5.2 2.3 virginica  146 6.3 2.5 5.0 1.9 virginica  147 6.5 3.0 5.2 2.0 virginica  148 6.2 3.4 5.4 2.3 virginica  149 5.9 3.0 5.1 1.8 virginica  [150 rows x 5 columns] |