|  |  |
| --- | --- |
| **Practicum Case** |  |
| COMP6140 | COMP6140001 | COMP6140049  Data Mining |
| **Computer Science** | **E241-COMP6140-RE01-09** |
| ***Valid on*** *Even Semester Year 2023/2024* | **Revision 00** |

## Learning Outcomes

* LO 2 – apply various data mining techniques

## Topic

* Session 09 – Classification I

## Sub Topics

* K Nearest Neighbors (k-NN)
* Decision Tree (ID3)

## Soal

*Case*

**Obesity Classification**

You are given the data for named **obesity.csv**. Do classification with **K-NN** and **Decision Tree** to classify the label of the individual. You must make sure that the algorithm already achieves the **highest** performance of the model. The details of the data are:

|  |  |  |
| --- | --- | --- |
| **Attribute** | **Type** | **Description** |
| ID | Number | A unique identifier for each individual |
| Age | Number | The age of the individual |
| Gender | Polynomial | The gender of the individual |
| Height | Number | The height of the individual in centimeters |
| Weight | Number | The weight of the individual in kilograms |
| BMI | Number | The body mass index of the individual, calculated as weight divided by height squared |
| Label [Label] | Polynomial | The obesity classification of the individual |

The **numerical value** of the **Type (Label)** has the definition shown below:

* Underweight - 0
* Normal Weight - 1
* Overweight - 2
* Obese - 3

Your tasks for the **classification** are:

* **Preprocess** the given training data based on data requirements that described on the table above.
* To achieve the **highest** performance, you must do some tasks below:
  + Select the **attributes** that want to be used for classification using **correlation weight.**
  + **Filter out attributes** which have lower correlation weight than average.
  + Apply the classification algorithm **K-NN** and **Decision Tree.**
  + Ensure that the selected algorithm must achieve the **highest** **accuracy**.
* Create a **model** from the **data** using the **algorithm** with the highest accuracy that you have find out in **previous step.**

**References:**

<https://www.kaggle.com/datasets/sujithmandala/obesity-classification-dataset>