|  |  |
| --- | --- |
| **Practicum Case** |  |
| COMP6140 | COMP6140001 | COMP6140049  Data Mining |
| **Computer Science** | **E241-COMP6140-RE01-07** |
| ***Valid on*** *Even Semester Year 2023/2024* | **Revision 00** |

## Learning Outcomes

* LO 2 – apply various data mining techniques

## Topic

* Session 07 – Clustering I

## Sub Topics

* Similarity Measurement
* K-Means

## Soal

*Case*

**Wine Clustering**

The datasets given consists of wine’s informations. From the datasets given, cluster to find out the information of the wine.

You are given the wine’s informations to be clustered in **wine.csv**. Do the clustering with **K-Means**. The **details** of the data are:

|  |  |  |
| --- | --- | --- |
| **Attribute** | **Type** | **Description** |
| Alcohol | Number | The alcohol of the wine |
| Malic\_Acid | Number | The malic acid of the wine |
| Ash | Number | The ash of the wine |
| Ash\_Alcanity | Number | The alcalinity of ash in the wine |
| Magnesium | Number | The magnesium of the wine |
| Total phenols | Number | The total phenols of the wine |
| Flavanoids | Number | The flavanoids of the wine |
| Nonflavanoid\_Phenols | Number | The nonflavanoid phenols of the wine |
| Proanthocyanins | Number | The proanthocyanins of the wine |
| Color\_Intensity | Number | The color intensity of the wine |
| Hue | Number | The hue of the wine |
| OD280 | Number | The OD280 of the wine |
| Proline | Number | The proline of the wine |

Your tasks for the **clustering** are:

* **Preprocess** the given data based on data requirements that described on the table above.
* **Normalize** the given data using **range transformation**.
* To achieve the **highest** **accuracy**, you must do some **tasks** below:
  + **Select** the **attributes** that want to be used for **clustering** using **correlation matrix.**
  + **Remove attributes**, that is correlated highly with other attributes.
  + **Apply** the clustering algorithm **K-Means.**
* Change every value of the **clustering** **result** based on the following detail:
  + **Cluster A**, for every wine which is in the cluster that have the following criteria:
* Have **lower alcohol**.
* Have **higher malic acid**.
* Have **higher ash**.
* Have **higher ash calamity**.
* Have **lower magnesium**.
* Have **lower total phenols**.
* Have **lower flavanoids**.
* Have **higher nonflavanoid phenols**.
* Have **lower proanthocyanins**.
* Have **higher color intensity**.
* Have **lower hue**.
* Have **lower OD280**.
* Have **lower proline**.
  + **Cluster B**, for every participant which is in the cluster that have the following criteria:
* Have **higher alcohol.**
* Have **lower malic acid.**
* Have **lower ash.**
* Have **lower ash calamity**.
* Have **higher magnesium**.
* Have **higher total phenols**.
* Have **higher flavanoids**.
* Have **lower nonflavanoid phenols**.
* Have **higher proanthocyanins**.
* Have **lower color intensity**.
* Have **higher hue**.
* Have **higher OD280**.
* Have **higher proline**.