# K-Nearest Neighbors

**Instructions:**

Please share your answers filled in-line in the word document. Submit code separately wherever applicable.

Please ensure you update all the details:

**Name: MD SABIULLAH Batch ID:**  PDS 07052024

**Topic: K-Nearest Neighbors**

**Guidelines:**

**1. An assignment submission is considered complete only when the correct and executable code(s) and documentation explaining the method and results are submitted. Failing to submit either of those will be considered an invalid submission and not a correct submission.**

**2. Ensure that you submit your assignments correctly and in full. Resubmission is not allowed.**

**3. Post the submission you can evaluate your work by referring to the keys provided. (will be available only post the submission).**

**Hints:**

1. **Business Problem**
   1. **What is the business objective?**
   2. **Are there any constraints?**
2. **Work on each feature of the dataset to create a data dictionary as displayed in the below image:**



**Make a table as shown above and provide information about the features such as its data type and its relevance to the model building. And if not relevant, provide reasons and a description of the feature.**

**Data Dictionary Of Glass :**

| **Column Name** | **Column Description** | **Data Type** | **Range** | **Min Value** | **Max Value** |
| --- | --- | --- | --- | --- | --- |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **RI** | **Refractive Index** | **Float** | **1.51115 - 1.53393** | **1.51115** | **1.53393** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Na** | **Sodium (weight percent in the glass)** | **Float** | **10.73 - 17.38** | **10.73** | **17.38** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Mg** | **Magnesium (weight percent)** | **Float** | **0 - 4.49** | **0** | **4.49** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Al** | **Aluminum (weight percent)** | **Float** | **0.29 - 3.5** | **0.29** | **3.5** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Si** | **Silicon (weight percent)** | **Float** | **69.81 - 75.41** | **69.81** | **75.41** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **K** | **Potassium (weight percent)** | **Float** | **0 - 6.21** | **0** | **6.21** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Ca** | **Calcium (weight percent)** | **Float** | **5.43 - 16.19** | **5.43** | **16.19** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Ba** | **Barium (weight percent)** | **Float** | **0 - 3.15** | **0** | **3.15** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Fe** | **Iron (weight percent)** | **Float** | **0 - 0.51** | **0** | **0.51** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Type** | **Glass type (categorical label)** | **Integer** | **1 - 7** | **1** | **7** |

1. **Data Pre-processing**

**3.1 Data Cleaning, Feature Engineering, etc.**

1. **Exploratory Data Analysis (EDA):**
   1. **Summary.**
   2. **Univariate analysis.**
   3. **Bivariate analysis.**
2. **Model Building**
   1. **Build the model on the scaled data (try multiple options).**
   2. **Perform KNN and use cross-validation techniques to get the optimum K value.**
   3. **Train and test the model and perform cross-validation techniques. Compare accuracies, precision, and recall and explain them in the documentation.**
   4. **Briefly explain the model output in the documentation.**
3. **Write about the benefits/impact of the solution - in what way does the business (client) benefit from the solution provided?**

**Problem Statement:**

1. A glass manufacturing plant uses different earth elements to design new glass materials based on customer requirements. For that, they would like to automate the process of classification as it’s a tedious job to manually classify them. Help the company achieve its objective by correctly classifying the glass type based on the other features using KNN algorithm.

A picture containing table

Description automatically generated