# 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:RAKESH SETHU NP Batch ID: 11052022\_7.30PM**

**Topic: K-Nearest Neighbors**

**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:**



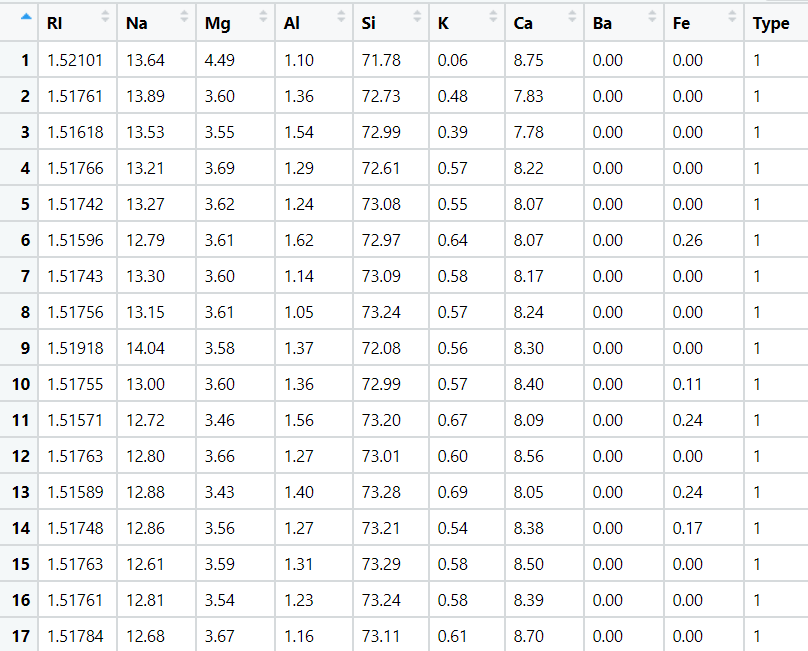
**2.1 Make a table as shown above and provide information about the features such as its data type and its relevance to the model building. And if not relevant, provide reasons and a description of the feature.**

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 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.**

1. **Write about the benefits/impact of the solution - in what way does the business (client) benefit from the solution provided?**
2. 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.



* 1. **What is the business objective?**

Predicting Glass Types

1. **Data Pre-processing**

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

Done data normalization

1. **Model Building**
   1. **Build the model on the scaled data (try multiple options).**
   2. **Perform KNN and use cross validation techniques to get 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.**

**1,Data split in X-train , y-train, x\_test and y\_test splitting for 80% train and 20% test**

1. **Preparing KNN model on train data with 7 neighbour then testing on test data**

**Accuracy score in test data is 86%**

**Crosstab**

**Predictions 1 2 3 5 7**

**Actual**

**1 10 2 1 0 0**

**2 7 13 0 0 0**

**3 4 1 0 0 0**

**5 0 0 0 2 1**

**6 0 1 0 0 0**

**7 0 0 0 0 1**

**Accuracy score in test data is 81%**

**Cross tab**

**Predictions 1 2 3 5 6 7**

**Actual**

**1 48 5 1 0 0 0**

**2 7 52 2 1 0 0**

**3 9 1 6 0 0 0**

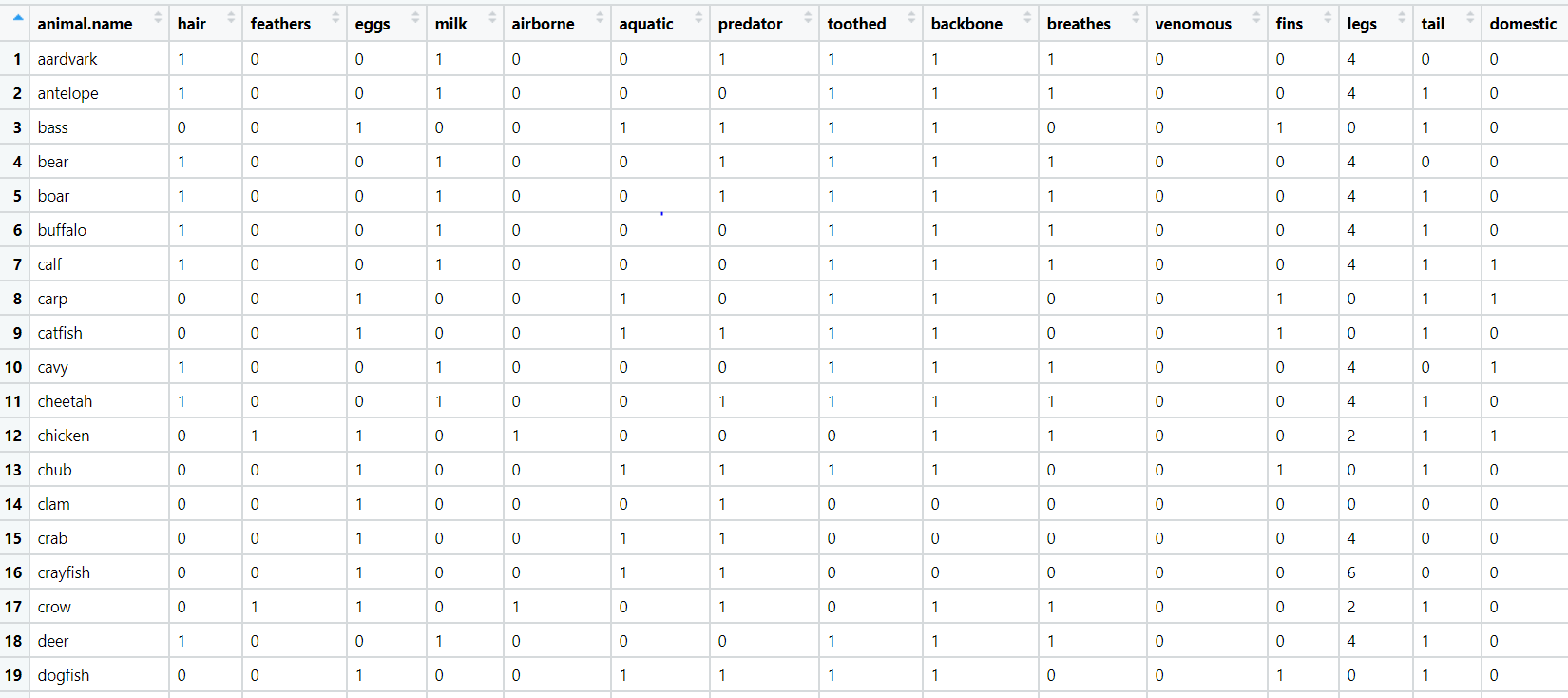
**5 0 1 0 10 0 1**

**6 0 1 0 0 6 0**

**7 1 0 0 1 1 17**

**And plotted them for right model**



1. A National Zoopark in India is dealing with the problem of segregation of the animals based on the different attributes they have. Build a KNN model to automatically classify the animals. Explain any inferences you draw in the documentation.
   1. **What is the business objective?**

Prediction of animal type

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

Done data normalization

1. **Model Building**
   1. **Build the model on the scaled data (try multiple options).**
   2. **Perform KNN and use cross validation techniques to get 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.**

**1,Data split in X-train , y-train, x\_test and y\_test splitting for 80% train and 20% test**

1. **Preparing KNN model on train data with 7 neighbour then testing on test data**

**Accuracy score in test data is 90%**

**Crosstab**

**Predictions 1 2 4 6 7**

**Actual**

**1 8 0 0 0 0**

**2 0 5 0 0 0**

**3 0 1 0 0 0**

**4 0 0 2 0 0**

**6 0 0 0 2 1**

**7 0 0 0 0 2**

**Accuracy score in test data is 97%**

**Cross tab**

**Predictions 1 2 3 4 5 6 7**

**Actual**

**1 33 0 0 0 0 0 0**

**2 0 15 0 0 0 0 0**

**3 0 0 4 0 0 0 0**

**4 0 0 0 11 0 0 0**

**5 0 0 0 0 4 0 0**

**6 0 0 0 0 0 4 1**

**7 0 0 1 0 0 0 7**

**And plotted them for right model**