**STANDARDIZATION & NORMALIZATION**

**NAME: ANANDAKRISHNAN K V**

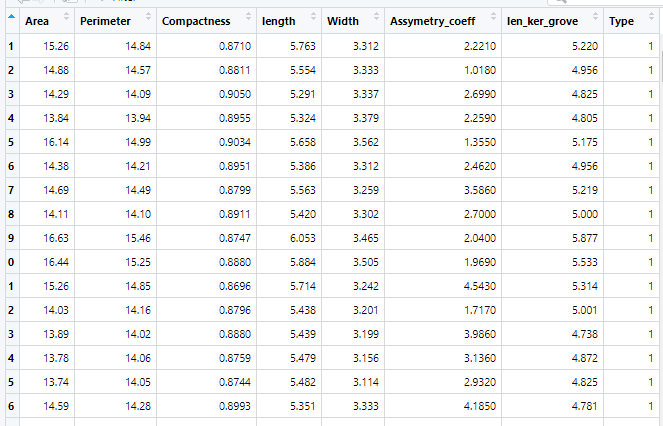
**BATCH ID: 19042021**

**Problem Statement:**

Data is one of the most important assets. It is often common that data is stored in distinct systems with different formats and scales. These seemingly small differences in how the data is stored can result in misinterpretations and inconsistencies in your analytics. Inconsistency can make it impossible to deliver reliable information to management for good decision making. We have the preprocessing techniques to make the data uniform. Explore the various techniques to have reliable uniform standard data, you can go through this link:

<https://360digitmg.com/mindmap-data-science>

1. Prepare the dataset by performing the preprocessing techniques, to have the standard scale to data which improves the model predictions .



**Ans:-**

**Python code:-**

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

from scipy import stats

from sklearn.impute import SimpleImputer as si

### import dataset Seeds\_data as dataframe ##

df = pd.DataFrame(Seeds\_datacsv)

df1 = df.copy(deep=True)

### Here the data set is not a mixed data set. Since all the datas are numeric for better visualisation normalization Scaling not suits here. Both normalization & standardization will work here. But since the output column "TYPE" is a ordinal data, thats even in the range 1 to 3 standardization scaling will gives visually better scaled output data set. So gonna apply standardisation scaling here

### standardization scaling ###

df.describe().T

#Importing the Libraries

from sklearn.preprocessing import StandardScaler, MinMaxScaler,RobustScaler

# define standard scaler

scaler = StandardScaler() # Standard Scaler or Standardization

# Transform data ## "TYPE" column is considered as ouput. so not gonna do any scaling there

df1.iloc[:,0:7] = scaler.fit\_transform(df1.iloc[:,0:7]) #Fit to data, then transform it.

print("Standardized Scaler :\n",df1)

df1.describe().T

### mean of each variable approximately = 0; stndered deviation of each variable approximately = 1. so satisfying standardization criteria

**Hints:**

For each assignment, the solution should be submitted in the below format

1. Work on each feature to create a data dictionary as displayed in the image displayed below:
2. Refer to Seeds\_data.csv file
3. Research and perform all possible steps for obtaining solution
4. All the codes (executable programs) should execute without errors
5. Code modularization should be followed
6. Each line of code should have comments explaining the logic and why you are using that function