**Zero - Variance Features**

Instruction

Please ensure you update all the details:

**Name: s**

**Batch Id:**

**Topic: Preliminaries for Data Analysis**

Variance measures how far a set of data is spread out. A variance of zero indicates that all the data values are identical. There are various techniques to remove this for transforming the data into the suitable one for prediction.

**Problem statement:**

Find which columns of the given dataset with zero variance, explore various techniques used to remove the zero variance from the dataset to perform certain analysis.



**Answer:**

|  |  |  |  |
| --- | --- | --- | --- |
| Name of Feature | Description | Types | Relevance |
| Id | Index no. | Quantitative,Discrete, Ordinal | Irrelevant |
| square.length | Squared length of data | Quantitative,Discrete, Ratio | Relevant |
| square.breadth | Squared breadth of data | Quantitative,Discrete, Ratio | Relevant |
| rec.Length | Recorded length of data | Quantitative,Discrete, Ratio | Relevant |
| rec.breadth | Recorded breadth of data | Quantitative,Discrete, Ratio | Relevant |
| colour | Color of the data | Qualitative,Discrete,  Nominal | Irrelevant |

**Objective:** To remove the zero variance from the dataset to perform certain analysis.

**Constraint:** To find the columns of the given dataset with zero variance.

**Working with the Data**

* In order to start working with the dataset, we need to import the data first and the necessary libraries
* We use either Spyder, Jupyter, Colab, as per your system requirement, ease of use and availability.

**Pre processing and EDA**:

* We check the data description using “info” or “describe”
* Additionally checking the shape shows that the data consists of 150 rows and 6 columns/ features
* We check the data for missing values using the function “isna()” and summing it up with “sum()”
* It describes that there are no missing values present in the data
* We removed the outliers using winsorization technique
* Here we used var() to know the variance of numerical variables
* Hence the data is now pre-processed and EDA performed. It is ready for further processing!

**Hints:**

A picture containing shape, arrow

Description automatically generatedFor each assignment, the solution should be submitted in the below format

1. Work on each feature of the dataset to create a data dictionary as displayed in the below image:



1. Consider the Z\_dataset.csv dataset
2. Research and perform all possible steps for obtaining solution
3. All the codes (executable programs) should execute without errors
4. Code modularization should be followed
5. Each line of code should have comments explaining the logic and why you are using that function