**Zero - Variance Features**

Instruction

Please ensure you update all the details:

**Name: DHEERAJ MISHRA**

**Batch Id: DS\_01072021**

**Topic: Data Pre-Processing**

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.



**SOLN:-**

1. **Business understanding:**

**Max:- Measurement techniques**

**Min:- Error in measurement**

**Constraints:- Lack of dataset**

1. **Data understanding**

|  |  |  |  |
| --- | --- | --- | --- |
| **Name of Feature** | **Description** | **Types** | **Relevance** |
| **Id** | **Sr number** | **Nominal , discrete** | **Id does not provide useful information** |
| **Square.length** | **Length of square** | **Continuous** | **Relevant** |
| **square.breadth** | **Breadth of square** | **Continuous** | **Relevant** |
| **rec.Length** | **Length of rectangle** | **Continuous** | **Relevant** |
| **rec.breadth** | **Breadth of rectangle** | **Continuous** | **Relevant** |
| **colour** | **Prediction** | **Categorical** | **Relevant** |

**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

**Grading Guidelines:**

**Note: 1. An Assignment submission is considered complete only when successful executable code(s), and documentation explaining the applied solution and results are provided. Failing to submit either of them will be considered an invalid submission and will not be considered for evaluation.**

**2. Assignments submitted after the deadline date will affect your grades.**

**Grading:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Ans** | **Date** |  |  | **Ans** | **Date** |
| Correct | On time | A | 100 |  |  |
| 80% & above | On time | B | 85 | Correct | Late |
| 50% & above | On time | C | 75 | 80% & above | Late |
| 50% & below | On time | D | 65 | 50% & above | Late |
|  |  | E | 55 | 50% & below |  |
| Copied/No Submission |  | F | 45 |  |  |

* **Grade A: (>= 90):** When all assignments are submitted on or before the given deadline date
* **Grade B: (>= 80 and < 90):** 
  + When assignments are submitted on time but less than 80% of questions asked in assignments are completed. (or)
  + All assignments were submitted, however, after the given deadline
* **A picture containing shape, arrow

  Description automatically generatedGrade C: (>= 70 and < 80):** 
  + When assignments are submitted on time but less than 50% of questions asked in assignments are completed. (or)
  + Less than 80% of questions asked in assignments are submitted after the deadline
* **Grade D: (>= 60 and < 70):** Assignments submitted after the Deadline and with 50% or less of questions
* **Grade E: (>= 50 and < 60):** 
  + Less than 30% of questions asked in the assignments are submitted after the deadline (OR)
  + Less than 30% of questions asked in the assignments are submitted before deadline

**Grade F: (< 50):** Copied submission or No submission