Introduction to Data Science

Midterm Project

Apply data preparation steps (which can be applied) and do the univariate data exploration for the given data set.

Data set description:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | **Caesarian Section Classification Dataset Data Set**  **Abstract**: This dataset contains information about caesarian section results of 80 pregnant women with the most important characteristics of delivery problems in the medical field. |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **Data Set Characteristics:** | Univariate | **Number of Instances:** | 80 | **Area:** | Life | | **Attribute Characteristics:** | Integer | **Number of Attributes:** | 5 | **Date Donated** | 2018-11-02 | | **Associated Tasks:** | Classification | **Missing Values?** | N/A | **Number of Web Hits:** | 71071 |   **Data Set Information:**  Provide all relevant information about your data set.  **Attribute Information:**  We choose age, delivery number, delivery time, blood pressure and heart status. We classify delivery time to Premature, Timely and Latecomer. As like the delivery time we consider blood pressure in three statuses of Low, Normal and High moods. Heart Problem is classified as apt and inept.  @attribute 'Age' { 22,26,28,27,32,36,33,23,20,29,25,37,24,18,30,40,31,19,21,35,17,38 } @attribute 'Delivery number' { 1,2,3,4 } @attribute 'Delivery time' { 0,1,2 } -> {0 = timely , 1 = premature , 2 = latecomer} @attribute 'Blood of Pressure' { 2,1,0 } -> {0 = low , 1 = normal , 2 = high } @attribute 'Heart Problem' { 1,0 } -> {0 = apt, 1 = inept }  @attribute Caesarian { 0,1 } -> {0 = No, 1 = Yes } |

**Instructions:**

1. Upload your Project ( R file or Text file) and Report to the Teams before deadline.
2. Submission Date**: March 13,2023**. (you must submit the assignment before **10:59 AM).**