**Birla Institute of Technology &amp; Science, Pilani**

**Work-Integrated Learning Program Division**

**M.Tech (Data Science and Engineering)**

**Dataset Name:** Chronic Kidney Disease

|  |  |
| --- | --- |
| **Abstract**: This dataset can be used to predict chronic kidney disease and it has been collected at a hospital for a period of nearly 2 months. |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Data Set Characteristics:** | Multivariate | **Number of Instances:** | 400 | **Area:** | N/A |
| **Attribute Characteristics:** | Real | **Number of Attributes:** | 25 | **Date Donated** | 2015-07-03 |
| **Associated Tasks:** | Classification | **Missing Values?** | Yes |  |  |

**Source:**

Source:   
Dr.P.Soundarapandian.M.D.,D.M   
(Senior Consultant Nephrologist),   
Apollo Hospitals,   
Managiri,   
Madurai Main Road,   
Karaikudi,   
Tamilnadu,   
India.

**Problem Statement**

Analyze the data set, investigate and evaluate the result and predict the overall performance.

The dataset has been uploaded to canvas.

The students need to

1. Select a method for performing the analytic task
2. Preprocess the data to enhance quality
3. Carry out descriptive summarization of data and make observations
4. Identify relevant, irrelevant attributes for building model.
5. Perform appropriate data transformations with justifications
6. Generate new features if needed
7. Carry out the chosen analytic task. Show results including intermediate results, as needed
8. Evaluate the solutions
9. Look for refinement opportunities

Following are some points for you to take note of, while doing the assignment:

* Prepare a Report that will supplement the submitted codebase.
* State all your assumptions clearly
* List all intermediate steps and learnings
* Make the report structured and readable.