After understanding the requirements,

* I will see on what platform the dataset can be made. Either excel, SQL, word, or any other format.
* What are the required fields in the dataset.
* On what grounds it shall be classified.
* Considering which type of variable I can include in the dataset.
* Understanding the limitations and capabilities of the dataset.
* Check if any row/column is empty.
* Dealing with key factors.
* Splitting the data into training (80%) and testing (20%) sets.
* Cross validation is an important thing to do.