1. CHOOSE DATASET
2. Name of Dataset: ionosphere
3. Number of Instances: 351
4. Number of Attributes: 34 plus the class attribute
5. If there are a number of NA or null values, how you plan to handle that situation- R is capable of handling missing values. In **R**, missing values are represented by the symbol **NA** (not available). **R** uses the same symbol for character and numeric data. Methods in R to handle data-
6. **is.na(x)** returns TRUE of x is missing
7. **complete.cases()** returns a logical vector indicating which cases are complete
8. **na.omit()** returns the object with list-wise deletion of missing values.
9. **na.fail**: returns the object only if it contains no missing values

Although there is no missing data or null values in this dataset.

1. The dataset consists of negative values. So we have normalized the dataset by making all the value between 0 and 1. For ease of use and data modelling, the target class values have been replaced. “g” is replaced with “1” and “b” is replaced with “0”.
2. UNDERSTAND THE DATASET
3. How was the data obtained (source of the data):

**Donor**: Vince Sigillito ([vgs@aplcen.apl.jhu.edu](mailto:vgs@aplcen.apl.jhu.edu))

**Date:** 1989

**Source:** Space Physics Group

Applied Physics Laboratory

Johns Hopkins University

1. What did the authors use the dataset for (use cases for the dataset): Used for Classification of radar returning from the ionosphere using neural networks.
2. What type of experiments were done on the dataset (experiments): They investigated using backpropagation and the perceptron training algorithm on this database
3. Summary of the authors’ results (this could be the accuracy that the authors obtained or other forms of results): Using the first 200 instances for training, which were carefully split almost 50% positive and 50% negative, they found that a "linear" perceptron attained 90.7%, a "non-linear" perceptron attained 92%, and backpropagation an average of over 96% accuracy on the remaining 150 test instances, consisting of 123 "good" and only 24 "bad" instances. Accuracy on "good" instances was much higher than for "bad" instances. Backprop was tested with several different numbers of hidden units (in [0,15]) and incremental results were also reported. David Aha briefly investigated this database. He found that nearest neighbor attains an accuracy of 92.1%, that Ross Quinlan's C4 algorithm attains 94.0% (no windowing), and that IB3 (Aha \& Kibler, IJCAI-1989) attained 96.7% (parameter settings: 70% and 80% for acceptance and dropping respectively).
4. Any other information you found interesting: This radar data was collected by a system in Goose Bay, Labrador. This system consists of a phased array of 16 high-frequency antennas with a total transmitted power on the order of 6.4 kilowatts. The targets were free electrons in the ionosphere. "Good" radar returns are those showing evidence of some type of structure in the ionosphere. "Bad" returns are those that do not; their signals pass through the ionosphere. Received signals were processed using an autocorrelation function whose arguments are the time of a pulse and the pulse number. There were 17 pulse numbers for the Goose Bay system. Instances in this database are described by 2 attributes per pulse number, corresponding to the complex values returned by the function resulting from the complex electromagnetic signal.