* Classification—Deciding if something belongs to one category or another
* Finding relations—Finding correlations or potential causes of effects seen in the data
* Characterization—Very general plotting and report generation from data
* Distribution – Too narrow Vs. Too wide, good predictor if wide, varies. You should rely on information about the problem domain to judge if the data range is narrow, but a rough rule of thumb is the ratio of the standard deviation to the mean. If that ratio is very small, then the data isn’t varying much.

Documentation

Steps were taken to ensure proper documentation and flawless reproducibility of the project from the start. For documentation, the codes were written in R Markdown file using the knitr package, which is basically a stream of texts and codes. This ensured that the documentation was synchronized with codes, results were in sync with data, and it provided efficient means of collaboration between the project contributors. The project history was tracked from the beginning using Git, the most popular version control system, and GitHub, the most popular code sharing platform based on Git. All the data were saved in non-proprietary format to ensure reproducibility.