Rules of Machine Learning

First design and implement metrics : It is followed by using the precipitation level for rain prediction using pressure, temperature data. First designing the system using these inputs and process using some metrics and implementing these metrics will predict whether it will rain tomorrow or not.

Keep the first model simple and get the infrastructure right.: This rule is being considered in the project while pre-classifying the data while loading data into BigQuery. Using the stored data and integrate that into the model.

Turn heuristics into features or handle them externally : Preprocessing the data to eliminate missing values, replacing missing values in the column with the median of the column values. Replacing the missing categorical data with most frequently used label.

Starting with an interpretable model makes debugging easier : The rule has been applied while predicting rain as probability which is the final output of our project. Using the logistical regression model which makes it easier to classify.

Plan to launch and iterate : We have discussed about future work on project that helps the end user to have an idea of percentage of rain that is the precipitation levels during that day. Training the data, preprocessing it and applying appropriate models to calculate the precipitation levels and output the percentage of rain during the day.

Clean up features you are no longer using : We have removed features like RainMM as they do not contribute in any way to the betterment of the model and instead can influence the output heavily.

If you produce a model based on the data until January 5th, test the model on the data from January 6th and after.: We will follow this rule and train the model till a certain day and try to make it predict on the subsequent days, to test if the performance of the model.