Data

The dataset was got from TRF and the historic yield was got from different plantations.

A table enlisting the variables and their characteristics is given below:

Sl. No.	Attribute	Range of available data	Description	Data Type
1	Temperature Min	Jan 1952-Dec 2020	Daily Minimum measured daily temperature	Float
2	Temperature Max	Jan 1952-Dec 2020	Daily Maximum measured daily temperature	Float
3	Humidity 8:30	Jan 1986-Dec 2020	Daily Humidity measured at 8:30 AM	Float
4	Humidity 2:30	Jan 1986-Dec 2020	Humidity measured at 2:30 PM	Float
5	Rainfall inches	Jan 1952-Dec 2020	Daily average Rainfall measured in inches	Float
6	Rainfall mm	Jan 1952-Dec 2020	Daily average Rainfall measured in mm	Float
7	Yield mm	Jan1980-Dec 2020	Monthly average yield measured in kg/ha	Float

This data has been collected and compiled through a combination of manual and automated methods. Rainfall has been measured at equal spatial intervals through a fibre-reinforcedplastic (FRP) rain-gauge. Rainfall for a specific period of time is calculated by averaging the measurement of water collected in the various rain-gauges present in the region. Measured rainfall of 1 mm, implies that the volume of water collected would be:
200 cm2 (dimension of the collection area) x /10cm = 20cm3, or 20 ml.

Further, Humidity is the measure of moisture content of the atmosphere, this is expressed as Relative Humidity which is the ratio of water vapour actually present in the atmosphere to the amount of water vapour required to saturate it at that temperature (expressed as a percentage). Relative humidity is measured using the wet and dry bulb thermometer readings and is read directly from hygrometric tables.

Tea yield is measured using the weight of Tea leaves plucked divided by the hectarage of the area harvested. Tea Yield has been captured in the dataset as monthly yield, which is the daily yield averaged over the course of a single month.

	Year	Month	Day	Temperature min	Temperature max	Humidity 8:30	Humidity 2:30	Rainfall Inches	Rainfall mm	dayofyear
1952-01-01	1952	1	1	5	23.888889	0.0	0.0	0.00	0.00	1
1952-01-02	1952	1	2	9.44444	23.888889	0.0	0.0	0.85	21.59	2
1952-01-03	1952	1	3	13.3333	25.000000	0.0	0.0	0.00	0.00	3
1952-01-04	1952	1	4	12.2222	22.777778	0.0	0.0	0.00	0.00	4
1952-01-05	1952	1	5	9.44444	22.777778	0.0	0.0	0.00	0.00	5

This is the dataset post preprocessing.