Introduction

Climate change and variability, as the defining predicament of the present moment has come to disrupt traditional practices across disciplines - with concerted efforts and radical approaches, now being prescribed to limit the potentially endless damage. Agriculture is one domain, facing such disruption on a planetary scale - crop culture owing to its inherent inter-linkage with weather patterns, is increasingly susceptible to the adverse impacts of climate change. This project is set to use machine learning to predict the crop yield with other climatic parameters. We find the location of the data and map it to the rainfall, humidity and daily temperature of the location.
This project aims to use several prominent machine learning algorithms such as Support Vector Machines, Support Vector Regression, Multi Layer Percepton, and K-Nearest Neighbours to quantify and evaluate the variation expressed in several climatic variables, and further assess the yield-response to this variation, of an economically noteworthy cash crop - Tea. The application of Machine Learning has enabled hitherto unprecedented breakthroughs in exposing underlying relationships between factors in numerous data-intensive fields such as image recognition and computer vision, natural language processing and control automation.

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 /10*cm* = 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.

Methodology:

For this Project, the
Python Programming language has been chosen owing to its extensive library eco-system, simple implementation, readability and flexibility. Many different Python Development environments exist with distinct features, for this project Jupyter Notebook has been selected as it is an end-end tool that enables data cleansing, pre-processing, visualization, training and testing of machine leaning models in an integrated environement. It supports a vast variety of packages, provides support for individual execution of independent code shells, and most of all provides ease of use without compromising functionality.
Several different packages have also been used to manipulate the data, train and fit models and also visualise the results. These packages are enlisted below:

- Pandas: For creating and manipulating dataframes.

- Folium: Python visualization library would be used to visualize the neighborhoods cluster distribution of using interactive leaflet map.

- Scikit Learn: For importing k-means clustering.
- Geocoder: To retrieve Location Data.
- Matplotlib: Python Plotting Module.

The algorithms used are:

- Support Vector Machine
- Support Vector Regression
- K-means Clustering
- MLP

Result:

Model to predict model was trained using two different sets of parameters

1) Parameters of model 1: Rainfall, Humidity and Temperature

   Efficiency of SVR: 25.77%

   Efficiency of MLP: 67.19%

2) Parameters of model 2: Rainfall, Humidity, Temperature and Historical yield

   Efficiency of SVR: 96.69%

   Efficiency of MLP: 94.93%

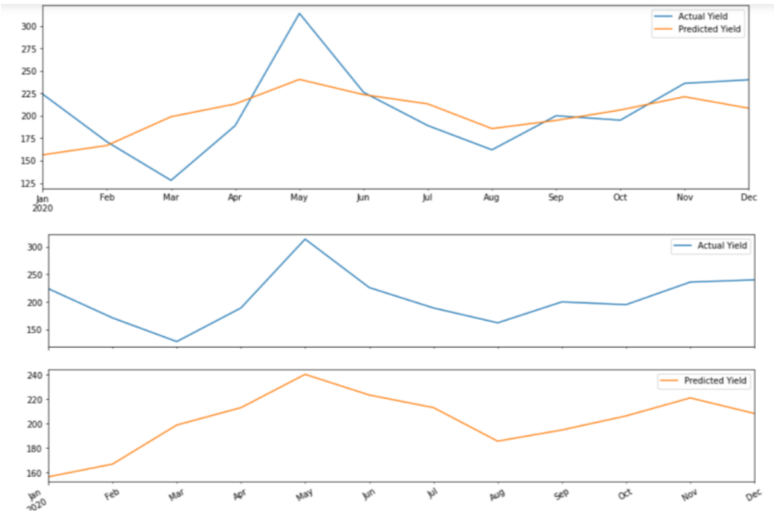
SVR:

	Actual Yield	Predicted Yield
2020-01-01	224.0	156.408011
2020-02-01	171.0	166.843335
2020-03-01	128.0	198.867367
2020-04-01	189.0	213.103956
2020-05-01	314.0	240.422385
2020-06-01	226.0	223.476042
2020-07-01	189.0	213.134619
2020-08-01	162.0	185.644642
2020-09-01	200.0	194.807987
2020-10-01	195.0	206.337181

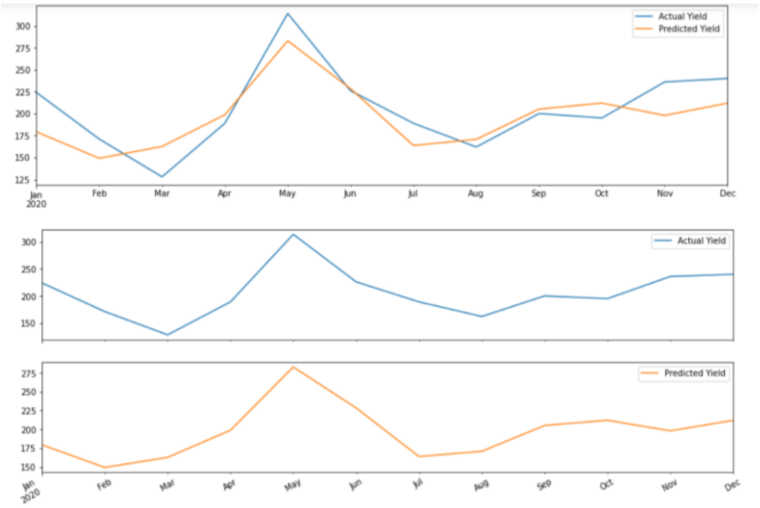
MLP:

	Actual Yield	Predicted Yield
2020-01-01	224	179.197661
2020-02-01	171	149.181311
2020-03-01	128	162.571922
2020-04-01	189	198.721428
2020-05-01	314	282.845096
2020-06-01	226	228.327316
2020-07-01	189	163.732088
2020-08-01	162	170.698845
2020-09-01	200	205.064914
2020-10-01	195	211.969376
2020-11-01	236	197.891108
2020-12-01	240	211.873551

The actual vs predicted values for SVR:



The actual vs predicted values for MLP:



Discussion:

The project finds the yield, and it is as follows, SVR and to a greater extent MLP was able to predict yield based on the weather data, to a considerable degree. The relatively lower degrees of accuracy when compared with the weather prediction from earlier could be attributed to the fact that several distinct factors exist that influence yield to varying degrees, therefore the variation in yield cannot be attributed entirely to weather factors.

Conclusion:

From this we can clearly see that there is a trend in yield data which grows with respect to the increasing trend in climatic parameters. And the yield quality and quantity varies with the condition of the area. A plantation 2000ft above sea level doesn’t give the same yield as the same cultivation done at 1000ft above sea level. Apart from the soil quality and the mineral rich environment, weather and other climatic conditions play a pivotal role in the cultivation.