Topic:-Crop Yield Prediction using Machine Learning.
Domain:-Machine Learning.

By-
Name:- Sneha Chavan.
Roll Number:-3220
Seminar Guide:- Prof. Y. A. Handge.

Index

- Introduction.
- Literature Survey.
- Motivation.
- Problem Definition.
- Software and Hardware Requirement.
- Implementation.
- Future Scope.
- Conclusion.
- References.

Introduction

- Weather conditions have a direct effect on crop yield. Various researches have been done exploring the connections between weather conditions and crop yield.

- The relations between crop yield and the weather, non-weather factors are non-linear.

- Algorithms like linear regression, decision tree and K-Nearest Neighbor (KNN) can help us to predict the crop yield.

Literature Survey

Index	Paper Name	Author Name	Conclusion
1.	Soybean Productivity Modelling using Decision Tree Algorithms	I. S. Veenadhari II. Dr. Bharat Mishrai III. Dr.CD Singh	Demonstrated how decision tree algorithm can increase the efficiency of yield prediction.
2.	Crop Prediction System using Machine Learning	I. Prof. D.S. Zingade II. Omkar Buchade III. Nilesh Mehta IV. Shubham Ghodekar V. Chandan Mehta	Demonstrated use of linear regression for yield prediction.
3.	Machine learning approach for forecasting crop yield	I. S.Veenadhari II. Dr. Bharat Misra III. Dr. CD Singh.	Demonstrated the use of data mining techniques for crop yield prediction.

Motivation

- Indian farmers may quit farming for the reasons for low productivity .
- To implement machine learning algorithms in agricultural field is important because it will help farmers to come up with better yield.
- It will help farmer to take decisions whether to take particular crop under certain conditions or not.

Problem Definition

Crop yield prediction based on the soil conditions and weather conditions using machine learning algorithms.

Keywords:-Machine Learning, prediciton.

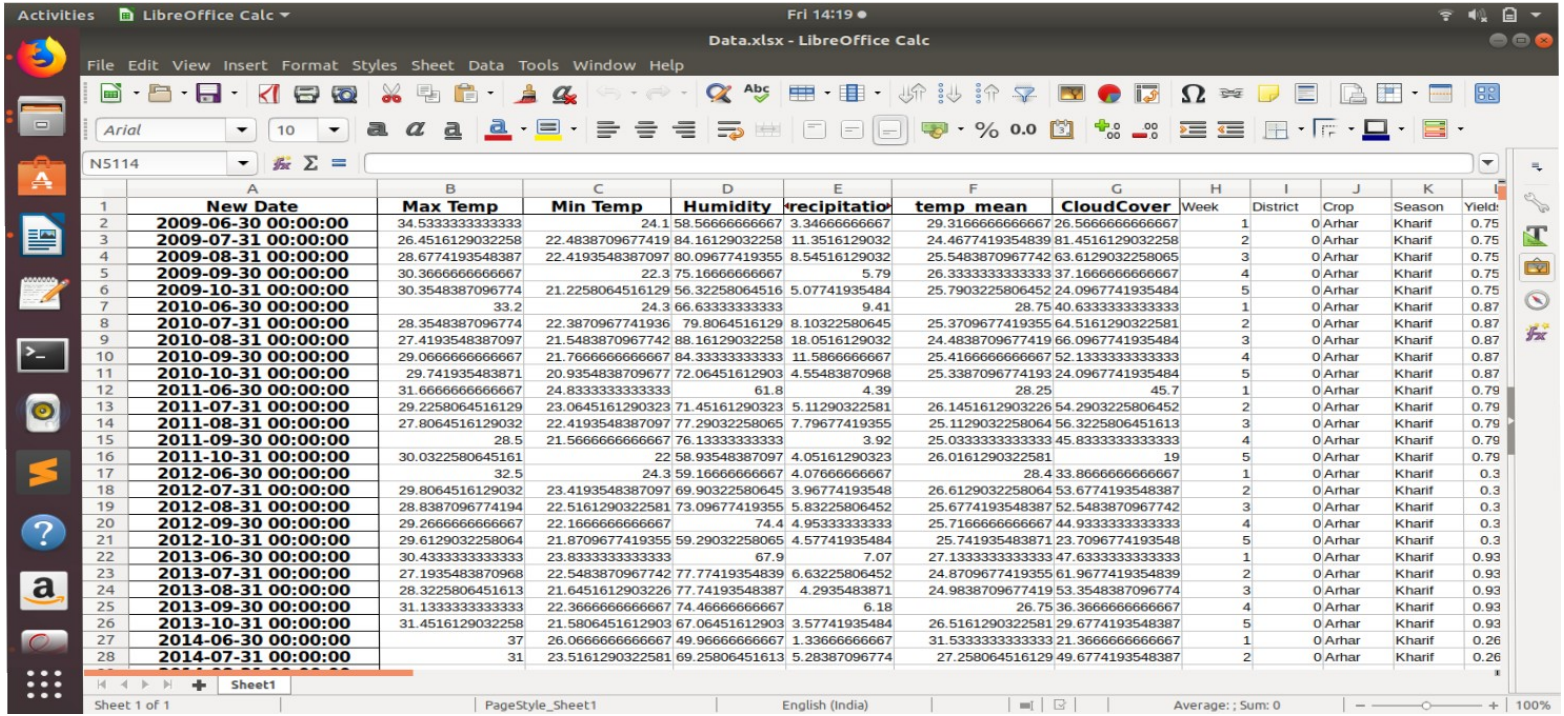
Software and Hardware Requirement

- Software Requirement:-
  I. Anaconda Navigator 3.0.1
  II.scikit-learn 0.20.3
  III.Numpy 1.15.4
  IV.Pandas 0.24.2
  V.Matplotlib 3.0.3

- Hardware Requirements:-
  I. Core i5 machine

Implementation

Dataset used



	A	B	C	D	E	F	G	H	I	J	K	L
1	New Date	Max Temp	Min Temp	Humidity	recipitatio	temp_mean	CloudCover	Week	District	Crop	Season	Yield
2	2009-06-30 00:00:00	34.5333333333333	24.1	58.56666666667	3.34666666667	29.3166666666667	26.5666666666667	1	0	Arhar	Kharif	0.75
3	2009-07-31 00:00:00	26.4516129032258	22.4838709677419	84.16129032258	11.3516129032	24.4677419354839	81.4516129032258	2	0	Arhar	Kharif	0.75
4	2009-08-31 00:00:00	28.6774193548387	22.4193548387097	80.09677419355	8.54516129032	25.5483870967742	63.6129032258065	3	0	Arhar	Kharif	0.75
5	2009-09-30 00:00:00	30.3666666666667	22.3	75.16666666667	5.79	26.3333333333333	37.1666666666667	4	0	Arhar	Kharif	0.75
6	2009-10-31 00:00:00	30.3548387096774	21.2258064516129	56.32258064516	5.07741935484	25.7903225806452	24.0967741935484	5	0	Arhar	Kharif	0.75
7	2010-06-30 00:00:00	33.2	24.3	66.63333333333	9.41	28.75	40.6333333333333	1	0	Arhar	Kharif	0.87
8	2010-07-31 00:00:00	28.3548387096774	22.3870967741936	79.8064516129	8.10322580645	25.3709677419355	64.5161290322581	2	0	Arhar	Kharif	0.87
9	2010-08-31 00:00:00	27.4193548387097	21.5483870967742	88.16129032258	18.0516129032	24.4838709677419	66.0967741935484	3	0	Arhar	Kharif	0.87
10	2010-09-30 00:00:00	29.0666666666667	21.7666666666667	84.33333333333	11.5866666667	25.4166666666667	52.1333333333333	4	0	Arhar	Kharif	0.87
11	2010-10-31 00:00:00	29.741935483871	20.9354838709677	72.06451612903	4.55483870968	25.3387096774193	24.0967741935484	5	0	Arhar	Kharif	0.87
12	2011-06-30 00:00:00	31.6666666666667	24.8333333333333	61.8	4.39	28.25	45.7	1	0	Arhar	Kharif	0.79
13	2011-07-31 00:00:00	29.2258064516129	23.0645161290323	71.45161290323	5.11290322581	26.1451612903226	54.2903225806452	2	0	Arhar	Kharif	0.79
14	2011-08-31 00:00:00	27.8064516129032	22.4193548387097	77.29032258065	7.79677419355	25.1129032258064	56.3225806451613	3	0	Arhar	Kharif	0.79
15	2011-09-30 00:00:00	28.5	21.5666666666667	76.13333333333	3.92	25.0333333333333	45.8333333333333	4	0	Arhar	Kharif	0.79
16	2011-10-31 00:00:00	30.0322580645161	22	58.93548387097	4.05161290323	26.0161290322581	19	5	0	Arhar	Kharif	0.79
17	2012-06-30 00:00:00	32.5	24.3	59.16666666667	4.07666666667	28.4	33.8666666666667	1	0	Arhar	Kharif	0.3
18	2012-07-31 00:00:00	29.8064516129032	23.4193548387097	69.90322580645	3.96774193548	26.6129032258064	53.6774193548387	2	0	Arhar	Kharif	0.3
19	2012-08-31 00:00:00	28.8387096774194	22.5161290322581	73.09677419355	5.83225806452	25.6774193548387	52.5483870967742	3	0	Arhar	Kharif	0.3
20	2012-09-30 00:00:00	29.2666666666667	22.1666666666667	74.4	4.95333333333	25.7166666666667	44.9333333333333	4	0	Arhar	Kharif	0.3
21	2012-10-31 00:00:00	29.6129032258064	21.8709677419355	59.29032258065	4.57741935484	25.741935483871	23.7096774193548	5	0	Arhar	Kharif	0.3
22	2013-06-30 00:00:00	30.4333333333333	23.8333333333333	67.9	7.07	27.1333333333333	47.6333333333333	1	0	Arhar	Kharif	0.93
23	2013-07-31 00:00:00	27.1935483870968	22.5483870967742	77.77419354839	6.63225806452	24.8709677419355	61.9677419354839	2	0	Arhar	Kharif	0.93
24	2013-08-31 00:00:00	28.3225806451613	21.6451612903226	77.74193548387	4.2935483871	24.9838709677419	53.3548387096774	3	0	Arhar	Kharif	0.93
25	2013-09-30 00:00:00	31.1333333333333	22.3666666666667	74.46666666667	6.18	26.75	36.3666666666667	4	0	Arhar	Kharif	0.93
26	2013-10-31 00:00:00	31.4516129032258	21.5806451612903	67.06451612903	3.57741935484	26.5161290322581	29.6774193548387	5	0	Arhar	Kharif	0.93
27	2014-06-30 00:00:00	37	26.0666666666667	49.96666666667	1.33666666667	31.5333333333333	21.3666666666667	1	0	Arhar	Kharif	0.26
28	2014-07-31 00:00:00	31	23.5161290322581	69.25806451613	5.28387096774	27.258064516129	49.6774193548387	2	0	Arhar	Kharif	0.26

Data Pre-processing

- Transforming raw data into an understandable format.
- Real-world data is often incomplete, inconsistent, and/or lacking in certain behaviors or trends.
- The data consisted of null values, float as well as integer values.
- Pre-processing involved converting null values into 0 and integer values into float.
- Data was pre-processed using panda in-built libraries.

Algorithms Used

- **Regression Algorithms -**
  - Regression algorithms predict the output values based on input features from the data fed in the system.
  - For predicting the crop yield three regression algorithms were implemented

    1)Linear Regression

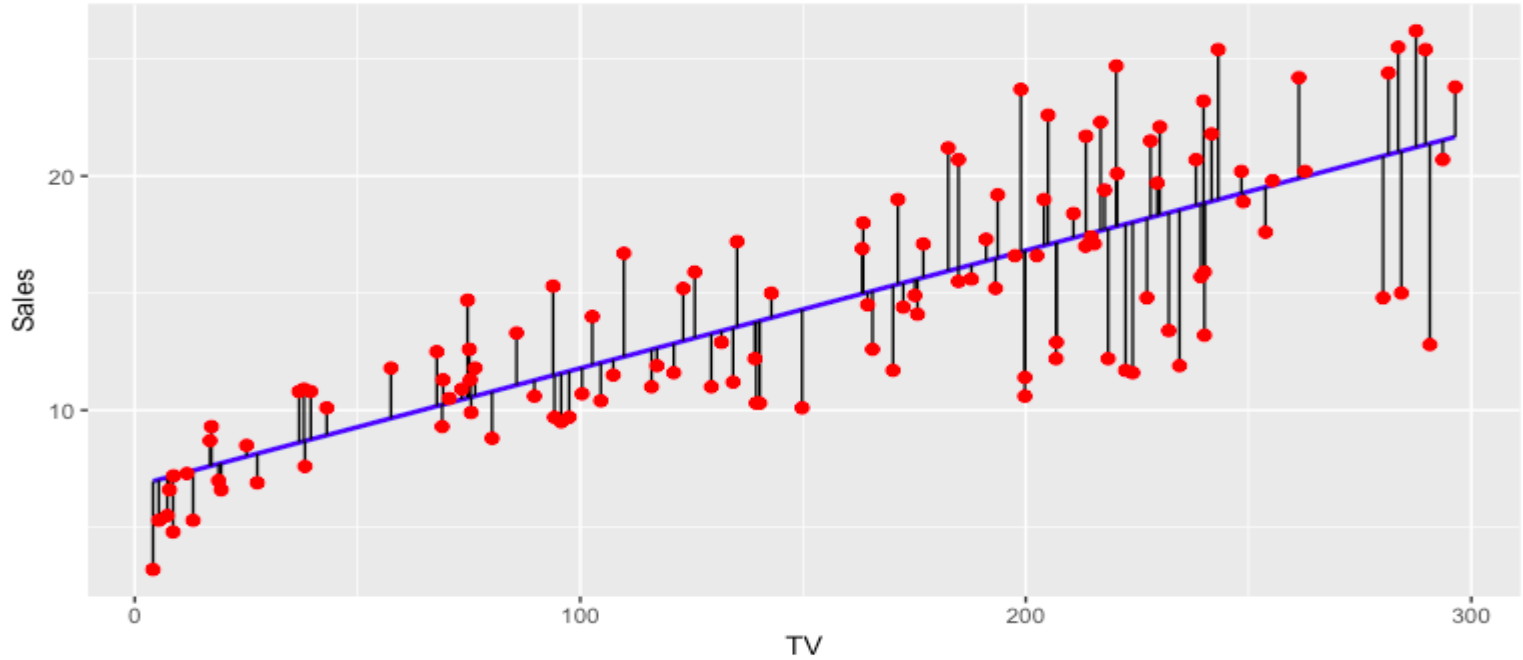
    2)Decision tree

    3)K-Nearest Neighbor (KNN)

Linear Regression

- Technique that is used to analyze response variable Y which changes with the value of the variable X.
- Y is known as dependent variable while X is independent variable.
- The least-square fit, which is capable of fitting both linear as well as polynomial relationships, is the most commonly used linear regression.

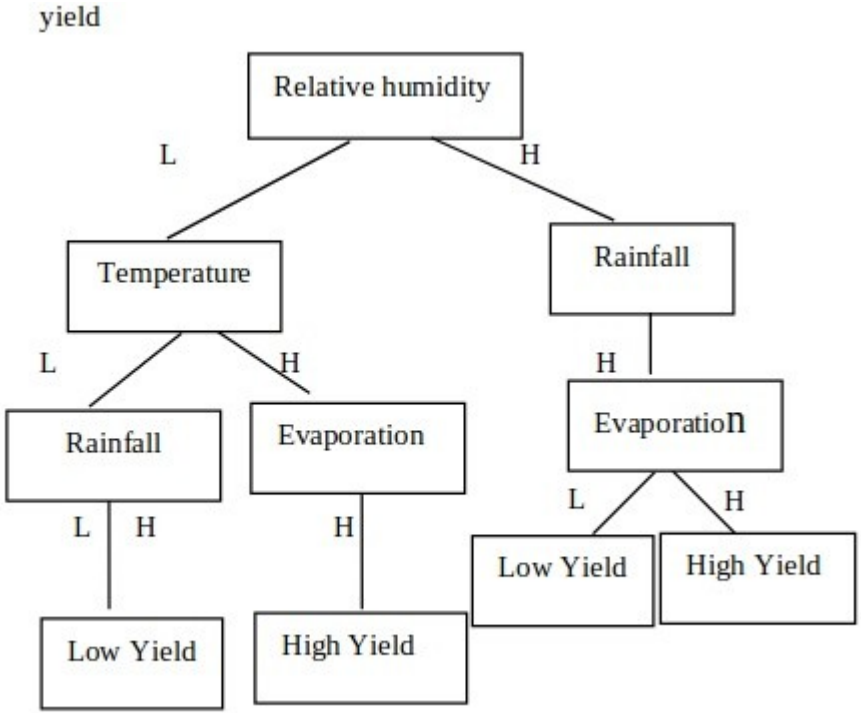
Basic Graph Of Linear Regression



Decision Tree

- The general motive of using Decision Tree is to create a training model which can use to predict class or value of target variables by learning decision rules inferred from prior data(training data).
- Steps in Decision Tree Algorithm
  - Place the best attribute of the dataset at the root of the tree.
  - Split the training set into subsets. Subsets should be made in such a way that each subset contains data with the same value for an attribute.
  - Repeat step 1 and step 2 on each subset until you find leaf nodes in all the branches of the tree.

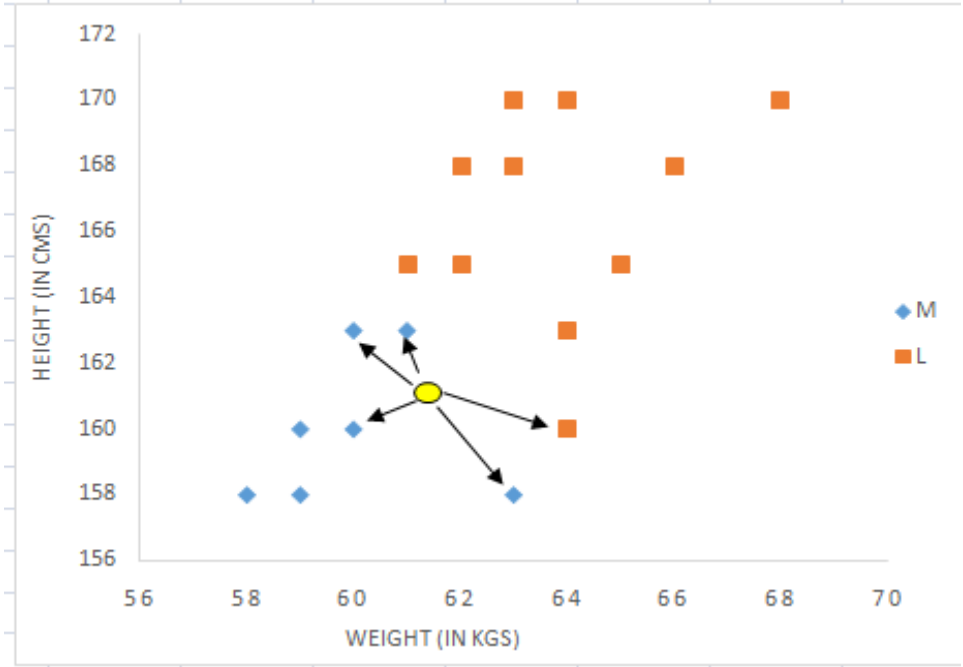
Decision Tree For Yield Prediction



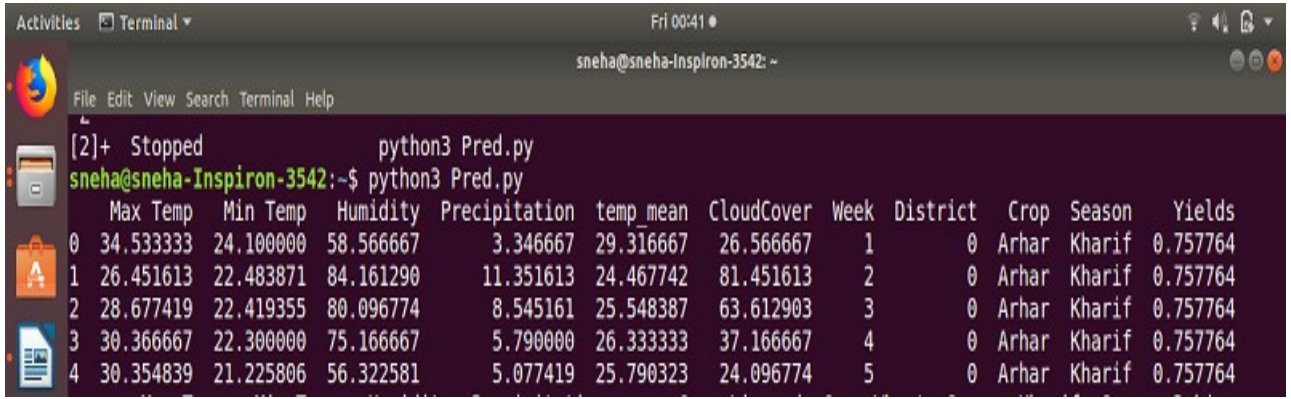
KNN

- In pattern recognition, the k-nearest neighbors algorithm (k-NN) is used for classification and regression.
- In both cases, the input consists of the k closest training examples in the feature space.
- The algorithm uses ‘feature similarity’ to predict values of any new data points. This means that the new point is assigned a value based on how closely it resembles the points in the training set.

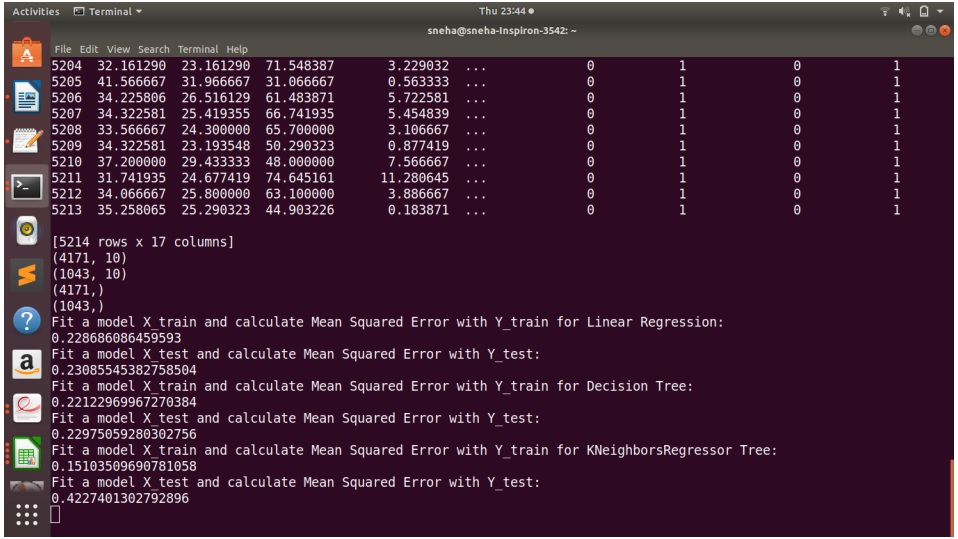
Working of KNN



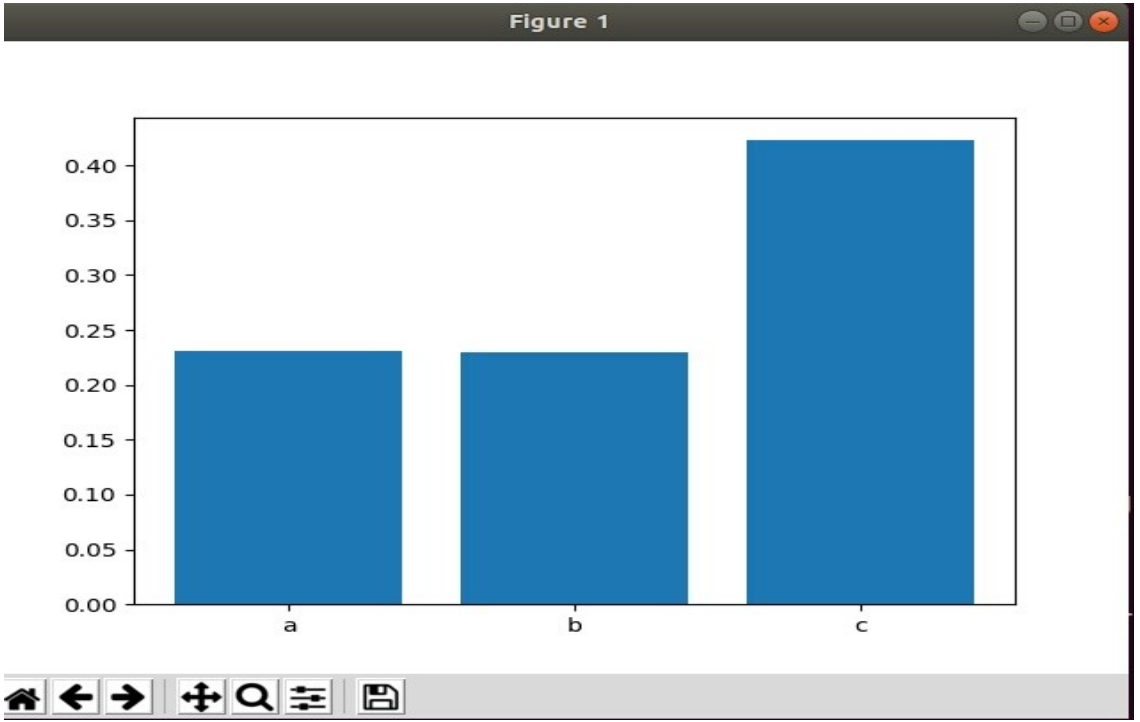
Screen shots from Implementation



Mean squared Error of three algorithms.



Graphical Representation of Error



Future Scope

The system can be enhanced further to add following functionality:

- Crop diseases detection using Image Processing where users can upload picture of diseased crop and get pesticides recommendations
- To develop fully automated monitoring system, which provides a real-time system that monitors soil parameters and weather conditions and will provide a real-time data time to time for analysis.
- Implementation of Smart Irrigation System to monitor weather and soil conditions, plant water usage etc. to automatically alter watering schedule.

Conclusion

- The system uses Machine learning algorithms and gives best result based on accuracy.
- The results of the three algorithms will be compared and the one giving the best and accurate output will be selected.
- Here Linear Regression Algorithm proves to be best for the dataset.
- Thus the system will help reduce the difficulties faced by the farmers by predicting the yield of the crop before sowing.

References

[1] Vaneesbeer Singh,Abid Sarwar, “Analysis of soil and prediction of crop yield(Rice) using Machine Learning approach” IJARCSE,vol. 5, Issue 8,2017.

[2] Monali Paul, Santosh K. Vishwakarma, Ashok Verma, “Analysis of Soil Behavior and Prediction of Crop Yield using Data Mining approach”, 2015 International Conference on Computational Intelligence and Communication Networks.

[3] S. Veenadhari, Dr. Bharat Mishra, Dr.CD Singh,”Soybean Productivity Modelling using Decision Tree Algorithms”,International Journal of Computer Applications (0975 –8887)Volume 27–No.7, August 2011

[4] Prof.D.S. Zingade1, Omkar Buchade2, Nilesh Mehta3, Shubham Ghodekar4, Chandan Mehta , ”Crop Prediction System using Machine Learning”, Dec.-2017,International Journal of Advance Engineering and Research Development