Machine Learning Approach For Crop Yield Prediction

Supervisor,
Dr. B S Harish

Team Members

1. Deepthi G N (01JST17IS015)
2. Madhurya Shankar (01JST17IS025)
3. Sinchana Hathwar (01JST17IS054)
4. Namratha M (01JST17IS058)

Problem Statement

Predicting the crop yield based on parameters like temperature, rainfall and other soil features (pH, N, P, K etc.) using machine learning approaches.

Introduction

- India is a country where agriculture and agriculture related industries are the major source of living for the people.

- Predicting the crop yield prior to its harvest can help the farmers and Government organizations

- Hypothetically it is all possible with the help of Machine Learning.

- Food is perhaps the most important aspect of life, without it, humankind can’t survive. Hence the estimation of food is very demanding. So, we attempt to forecast the yield of crops based on agriculture parameters

Literature Review

A Study on Various Data Mining Techniques for Crop Yield Prediction. (IEEE 2017)

Author: Yogesh Gandge, Sandhya
Methods Used: Classification Algorithm

- Crop yield prediction per acre with some recommendation.
- A unified approach where in all the factors affecting the crop yield can be utilized simultaneously for edicting the crop yield was not used.

Agricultural Production Output Prediction Using Supervised Machine Learning Techniques.(IEEE 2017)

Author: Md. Tahmid Shakoor, Karishma Rahman, Sumaiya Nasrin Rayta, Amitabha Chakrabarty
Methods Used: k-Nearest Neighbor, Decision Tree algorithm, ID3(Iterative Dichotomis) algorithm.

- Decision Tree Learning- ID3 algorithm gives a less value for percentage error than the KNN algorithm without omitting the outliers of the dataset.
- Research is limited to some fixed data-set

Effect of Temperature and Rainfall on Paddy Yield using Data Mining. (IEEE 2017)

Authors: Kuljit Kaur, Kanwalpreet Singh Attwal
Methods Used: Apriori Algorithm.
- Paddy yield was found to be high at low rainfall and low during high rainfall.
- Parameters such as Rainfall and Temperature was considered.

Improving Crop Productivity Through A Crop Recommendation System Using Ensembling Technique.(IEEE 2018)

Authors: Nidhi H Kulkarni, Dr. G N Srinivasan, Dr. B M Sagar, Dr.N K Cauvery
Methods Used: Random Forest, Naive Bayes, and Linear SVM.

- Recommends suitable crops for the farmers.
- Uses less amount of datasets which leads to less accurate results and takes more time processing data.

On the Performance of Temporal Stacking and Vegetation Indices for Detection and Estimation of Tobacco Crop (IEEE 2018)

Author: Waleed Khan, Nasru Minallah, Imran Ullah Khan, Zahid Wadud, Muhammad, Zeeshan, Suhail Yousaf And Abdul Baseer Qazi
Methods Used: Regression and SVM.

- Predict the yield of crop tabacco.
- Cant be applied for multiple crops and not suitable for multiple regions.

Multilevel Deep Learning Network for County-Level Corn Yield Estimation in the U.S. Corn Belt (IEEE2020)

Author: Jie Sun , Zulong Lai, Liping Di , Senior Member, IEEE, Ziheng Sun, Jianbin Tao, and Yonglin Shen
Methods Used: Convolutional neural network (CNN) or recurrent neural network (RNN).

- Predict the yield of crop corn.
- Uses the parameters related to corn and the datasets not suitable for other crops yield prediction.

Prediction of Crop Cultivation (IEEE 2019)

Authors : Neha Rale, Raxitkumar Solanki, Doina Bein, James Andro-Vasko, Wolfgang Bein.

Methods Used : Linear regression with polynomial features, and support-vector regression using a Radial Basis Function (RBF) kernel.

- Linear regression and support vector regression generates outputs graphically which is difficult to analyze.
- Not suitable in real time and uses small Data-set used for predictio**n.**

Crop Yield Prediction and Efficient use of Fertilizers. (IEEE 2019)

Authors : S.Bhanumathi, M.Vineeth and N.Rohit.

Methods Used : Random Forest and Back propagation algorithm used for implementation.

- Less parameters used for yield prediction
- Based on fertilizers , system predicts crop yield , but not considering all agriculture parameters.

Crop Yield Prediction Using Deep Reinforcement Learning Model for Sustainable Agrarian Applications. (IEEE 2020)

Authors: Dhivya Elavarasan , P. M. Durairaj Vincent.

Methods Used: Constructs a Deep Recurrent Q-Network model which is a Recurrent Neural Network deep learning algorithm.

- Uses neural network techniques.
- Huge data required.
- More time required for prediction.

Supervised Machine learning Approach for Crop Yield Prediction in Agriculture Sector. (IEEE 2020)

Authors : Dr. Y. Jeevan Nagendra Kumar, V. Spandana, V.S. Vaishnavi, K. Neha.

Methods Used :Random Forest Machine learning Algorithm.

- Takes More time required for prediction.
- Not suitable for real time.
- Less accurate results.

Challenges of Existing System

- Manual processing
- Less Reliable and efficient
- Time Consuming
- No Automation

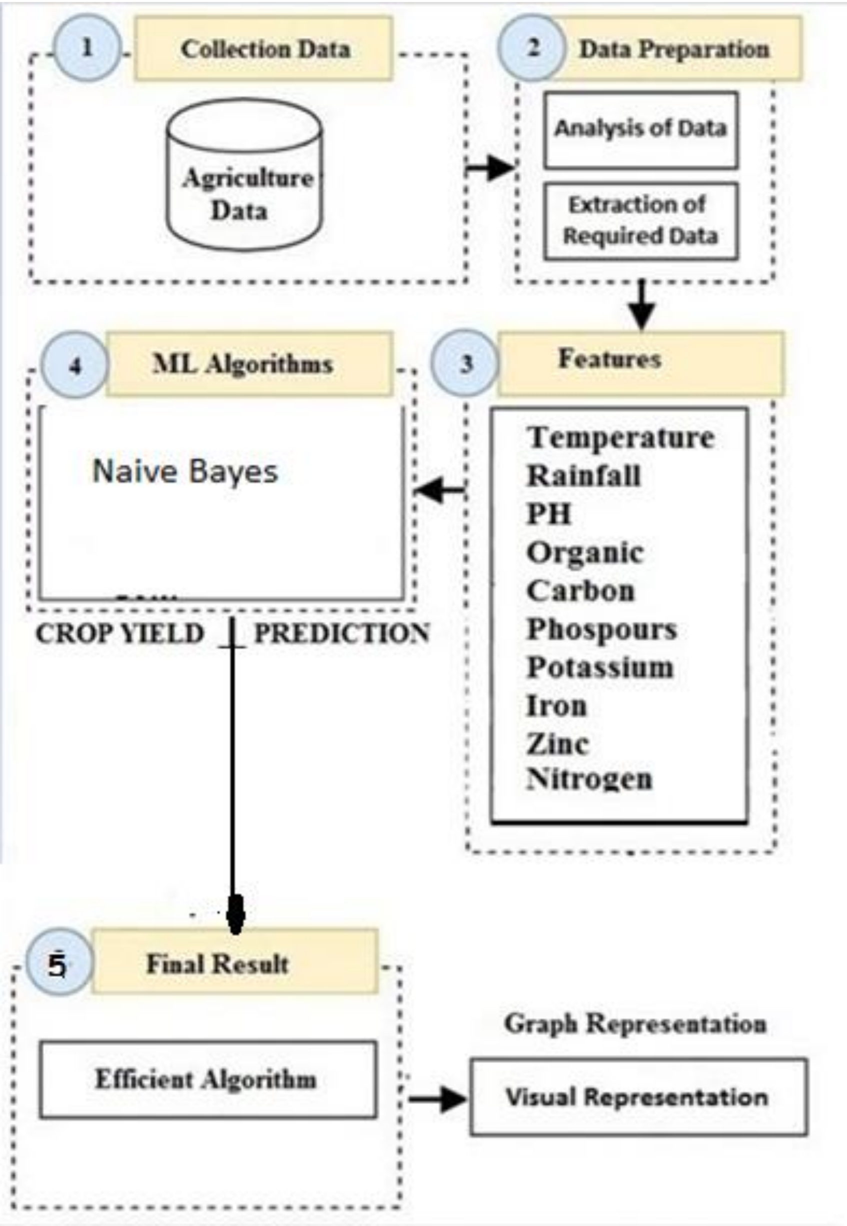
Motivation

- Agriculture is one of the most important occupation practiced in our country people.
- The Crop yield has a direct impact on National and International economies annually.
- Due to financial loss many farmers are committing suicide.
- Modernization of agriculture is very important and will lead the farmers of our country towards profit.

Objectives

- To develop a model by considering various parameters like temperature, rainfall, PH value, nitrogen, potassium, iron, zinc etc.
- To use Machine Learning algorithms to classify and predict the crop yield.
- To validate the proposed method by using the datasets.
- To test the developed model in real time.

Methodology



- Step 1: Collection Data

This is the first step in the Crop yield prediction process, where we collect agriculture data. We have collected the datasets for the crop Paddy and Ragi from agriculture department for the region Mysuru.

- Step 2: Data Preparation

Agriculture data is analyzed and only relevant data are extracted. We have used binning method for preprocessing.

- Step 3: Features

Agriculture parameters used for crop yield prediction are fetched.

- Step 4: ML Algorithms

In Supervised learning, you train the machine using data which is well labelled. We use Supervised Learning algorithm such as the Naive Bayes Algorithm for crop yield prediction.

- Step 5: Final Result

Crop yield is displayed for the farmers on GUI.

System Requirements

Software requirements

- OS : Windows 2000
- Back End : SQLSERVER
- Designed Tool Kit : Visual Studio 2010
- Front End : ASP.NET 4.0
- Programming Language : C#

Hardware requirements

- Intel P4 +
- 2 GB RAM +

Functional Requirements

- System is browser based application which predicts Rice and Ragi yield based on the soil test results and temperature, rainfall.
- System makes use of naive bayes algorithm for rice yield prediction.
- System generates accurate results based on the size of the dataset.

Non-Functional requirements

- Usability
- Reliability
- Maintainability
- Quality of service

Naïve Bayes

Sample example for Naïve Bayes :

Parameters – Temp, Humidity ,Area [m=3]
Outcome – 100T, 200T [p=1/2=0.5]

Training Dataset

Soil Type	Temp(L,M,H)	Humidity(L, M,H)	Area(100,200, 300)	Yield(subject)
2013	L	L	100	100T
2014	M	L	200	100T
2015	L	M	100	100T
2016	M	L	300	200T
2017	H	M	200	100T

2019 Parameters – temp- L, humidity - M, area - 200 Yield- ?

P=[n_c + (m*p)]/(n+m)

M P=[n_c + (m*p)]/(n+m) n=2, n_c=2,m=3,p=0.5 p=[2+(3*0.5)]/(2+3) p=0.7	M P=[n_c + (m*p)]/(n+m) n=2, n_c=0,m=3,p=0.5 p=[0+(3*0.5)]/(2+3) p=0.3
L P=[n_c + (m*p)]/(n+m) n=2, n_c=1,m=3,p=0.5 p=[1+(3*0.5)]/(2+3) p=0.5	L P=[n_c + (m*p)]/(n+m) n=2, n_c=1,m=3,p=0.5 p=[1+(3*0.5)]/(2+3) p=0.5
200 P=[n_c + (m*p)]/(n+m) n=2, n_c=2,m=3,p=0.5 p=[2+(3*0.5)]/(2+3) p=0.7	200 P=[n_c + (m*p)]/(n+m) n=2, n_c=0,m=3,p=0.5 p=[0+(3*0.5)]/(2+3) p=0.3

100T = 0.7 * 0.7 * 0.5 * 0.5 (p)
=0.1225

200T = 0.5 * 0.3 * 0.3 * 0.5 (p)
=0.0225

Since 0.1225 >0.0225
So this new Soil Input is classified to 100T

Why Naïve Bayes?

- It is highly scalable with the number of predictors and data points.
- It doesn't require as much training data.
- It is fast and can be used to make real-time predictions.

Project Overview

- **Input –** Previous year's agriculture data which includes temperature, rainfall, humidity and other constraints.

- **Output –** predicts crop yield using different constraints such as region, temperature, rain, humidity, soil features etc. and yield prediction based on year wise and location wise.

Snapshots of dataset

Rice

PH	nic carbon	nitrogen(n	osphorus(	otassium(k	sulphur(s)	zinc(zn)	iron(fe)	emperatur	Rainfall	Year	Crop	Yield
5.7	7	188.43	7	7.15	176	127.4	5.7	26.12	68.21	2012	Paddy(rice)	15
5.78	7.6	190.1	7.6	6.4	160.2	126.3	5.78	25.28	105.43	2013	Paddy(rice)	10
5.83	7.24	240.6	7.24	6.2	140	123	5.83	25.75	105.14	2014	Paddy(rice)	15
5.9	7.15	184.4	7.15	6	126.2	128.6	5.9	26.66	88.69	2015	Paddy(rice)	10
5.8	6.4	177	6.4	5.9	124.4	124.8	5.8	26.07	41.71	2016	Paddy(rice)	25
5.76	6.2	161.2	6.2	5.7	120.6	127.6	5.76	25.95	140.6	2017	Paddy(rice)	24
7.2	6	127.4	6	6.3	110	158	7.2	26.12	68.21	2012	Paddy(rice)	20
7.4	5.9	126.3	5.9	6.4	177.8	126.3	7.4	25.28	105.43	2013	Paddy(rice)	18
7	5.7	123	5.7	6.2	127	123	7	25.75	105.14	2014	Paddy(rice)	16
7.6	6.3	128.6	6.3	7	116.4	128.6	7.6	26.66	88.69	2015	Paddy(rice)	15
7.24	6.4	124.8	6.4	7.6	105.55	124.8	7.24	26.07	41.71	2016	Paddy(rice)	28
7.15	6.2	127.6	6.2	7.24	122.4	127.6	7.15	25.95	140.6	2017	Paddy(rice)	28
6.4	7	158	7	7.15	191.2	158	6.4	26.12	68.21	2012	Paddy(rice)	18
6.2	7.6	143.3	7.6	6.4	143.4	143.3	6.2	25.28	105.43	2013	Paddy(rice)	19
6	7.24	137.4	7.24	6.2	140.2	137.4	6	25.75	105.14	2014	Paddy(rice)	12
5.9	7.15	122	7.15	6	116.4	122	5.9	26.66	88.69	2015	Paddy(rice)	10
5.7	6.4	124.1	6.4	6.2	105.55	124.1	5.7	26.07	41.71	2016	Paddy(rice)	8
6.3	6.2	125.6	6.2	6	122.4	125.6	6.3	25.95	140.6	2017	Paddy(rice)	12
6.4	6	110	6	5.9	191.2	110	6.4	26.12	68.21	2012	Paddy(rice)	15
6.2	5.9	161.2	5.9	5.7	143.4	161.2	6.2	25.28	105.43	2013	Paddy(rice)	10
6.2	5.7	184.4	5.7	6.3	140.2	184.4	6.2	25.75	105.14	2014	Paddy(rice)	15
6.1	6.3	155	6.3	6.4	155.8	155	6.1	26.66	88.69	2015	Paddy(rice)	10
5.9	6.4	161.2	6.4	6.2	154.2	161.2	5.9	26.07	41.71	2016	Paddy(rice)	25
6.2	6.2	182	6.2	6.2	161.4	182	6.2	25.95	140.6	2017	Paddy(rice)	24
7.1	6.2	108	6.2	6.4	85	108	7.1	26.12	68.21	2012	Paddy(rice)	20
7.1	6.4	126.3	6.4	6.2	77.3	126.3	7.1	25.28	105.43	2013	Paddy(rice)	18
7.1	6.2	124.1	6.2	6.2	93.4	124.1	7.1	25.75	105.14	2014	Paddy(rice)	16
6.9	6.2	126.3	6.2	7	139.8	126.3	6.9	26.66	88.69	2015	Paddy(rice)	15
6.7	7	110	7	7.6	74.51	110	6.7	26.07	41.71	2016	Paddy(rice)	28

Ragi

	Name	PH	organic ca	nitrogen(n	phosphor	potassium	sulphur(s)	zinc(zn)	iron(fe)	Temperatu	Rainfall	Region	Year	Crop	Yield
2	Mahadev	7.4	5.9	126.3	5.9	6.4	177.8	126.3	7.4	25.28	105.43	Nagarle	2013	Wheat	18
3	Mahadeva	7	5.7	123	5.7	6.2	127	123	7	25.75	105.14	Nagarle	2014	Wheat	16
4	Narayana	7.6	6.3	128.6	6.3	7	116.4	128.6	7.6	26.66	88.69	Nagarle	2015	Wheat	15
5	Mahadeva	7.24	6.4	124.8	6.4	7.6	105.55	124.8	7.24	26.07	41.71	Nagarle	2016	Wheat	28
6	Ramesh	7.15	6.2	127.6	6.2	7.24	122.4	127.6	7.15	25.95	140.6	Nagarle	2017	Wheat	28
7	Siddaraju	6.4	7	158	7	7.15	191.2	158	6.4	26.12	68.21	Nagarle	2012	Wheat	18
8	Ramappa	6.2	7.6	143.3	7.6	6.4	143.4	143.3	6.2	25.28	105.43	Nagarle	2013	Wheat	19
9	Siddaraju	6	7.24	137.4	7.24	6.2	140.2	137.4	6	25.75	105.14	Nagarle	2014	Wheat	12
10	Gowdru	5.9	7.15	122	7.15	6	116.4	122	5.9	26.66	88.69	Nagarle	2015	Wheat	10
11	Ramappa	5.9	7.15	184.4	7.15	6	126.2	128.6	5.9	26.66	88.69	Belagunda	2015	Wheat	10
12	Nangunda	5.8	6.4	177	6.4	5.9	124.4	124.8	5.8	26.07	41.71	Belagunda	2016	Wheat	25
13	Prakash	5.76	6.2	161.2	6.2	5.7	120.6	127.6	5.76	25.95	140.6	Belagunda	2017	Wheat	24
14	Suresh	7.2	6	127.4	6	6.3	110	158	7.2	26.12	68.21	Nagarle	2012	Wheat	20
15	Mahadev	7.4	5.9	126.3	5.9	6.4	177.8	126.3	7.4	25.28	105.43	Nagarle	2013	Wheat	18
16	Mahadeva	7	5.7	123	5.7	6.2	127	123	7	25.75	105.14	Nagarle	2014	Wheat	16
17	Narayana	7.6	6.3	128.6	6.3	7	116.4	128.6	7.6	26.66	88.69	Nagarle	2015	Wheat	15
18	Mahadeva	7.24	6.4	124.8	6.4	7.6	105.55	124.8	7.24	26.07	41.71	Nagarle	2016	Wheat	28
19	Ramesh	7.15	6.2	127.6	6.2	7.24	122.4	127.6	7.15	25.95	140.6	Nagarle	2017	Wheat	28
20	Siddaraju	6.4	7	158	7	7.15	191.2	158	6.4	26.12	68.21	Nagarle	2012	Wheat	18
21	Ramappa	6.2	7.6	143.3	7.6	6.4	143.4	143.3	6.2	25.28	105.43	Nagarle	2013	Wheat	19
22	Siddaraju	6	7.24	137.4	7.24	6.2	140.2	137.4	6	25.75	105.14	Nagarle	2014	Wheat	12
23	Gowdru	5.9	7.15	122	7.15	6	116.4	122	5.9	26.66	88.69	Nagarle	2015	Wheat	10
24	Amir	6.73	6.2	38.5	6.2	6.4	155.8	38.5	6.73	25.95	140.6	Belagunda	2017	Wheat	10
25	Cheluvaraj	7.1	7	131.3	7	6.2	154.2	131.3	7.1	26.12	68.21	Haniyamb	2012	Wheat	8
26	raja	7.21	7.6	143.2	7.6	6.2	161.4	143.2	7.21	25.28	105.43	Haniyamb	2013	Wheat	12
27	Cheluvaraj	7.24	7.24	155.3	7.24	7	85	155.3	7.24	25.75	105.14	Haniyamb	2014	Wheat	15
28	raju	7.13	7.15	165.1	7.15	7.6	77.3	165.1	7.13	26.66	88.69	Haniyamb	2015	Wheat	10

Naïve Bayes Algorithm Steps

- **Step 1**: Scan the dataset

```
public ArrayList GetSubject()
{
        ArrayList s = new ArrayList();
        if (dtDistinct.Rows.Count > 0){
                for (int i = 0; i < dtDistinct.Rows.Count; i++){
                s.Add(dtDistinct.Rows[i]["Yield"].ToString());
        }
    }
return s;
}
```

- **Step 2**: Calculate the probability of each attribute value. [n, nc, m, p]

```
for (int i = 0; i < s.Count; i++)
    {
        mul.Clear();
        for (int j = 0; j < features.Length; j++)
        {
            n = 0;
            nc = 0;
            for (int d = 0; d < dt.Rows.Count; d++)
            {
                if (dt.Rows[d][j + 1].ToString().Equals(values[j]))
                {
                    ++n;
if (dt.Rows[d][m + 4].ToString().Equals(s[i]))
                            ++nc;
                    }
                }

                double x = m * p;
                double y = n + m;
                double z = nc + x;
                pi = z / y;
                mul.Add(Math.Abs(pi));
            }
```

- **Step 3**: Apply the formulae

P(attributevalue(ai)/subjectvalue(vj))=(nc + mp)/(n+m)

Where:

n = the number of training examples for which v = vj

nc = number of examples for which v = vj and a = ai

p = a prior estimate for P(ai/vj)

m = the equivalent sample size

- **Step 4**: Multiply the probabilities by for each class, here we multiple the results of each attribute with p and final results are used for classification.

```
for (int z = 0; z < mul.Count; z++)
        {
            mulres *= double.Parse(mul[z].ToString());
        }

        result = mulres * p;
        output.Add(Math.Abs(result));
    }
```

- **Step 5**: Compare the values and classify the attribute values to one of the predefined set of class.

```
for (int x = 0; x < s.Count; x++)
        {
            list1.Add(output[x]);
        }

        list1.Sort();
        list1.Reverse();

        string _output = null;

        for (int y = 0; y < s.Count; y++)
        {
            if (output[y].Equals(list1[0]))
            {
                _output = s[y].ToString();
                return _output;
            }
        }
```

Naïve Bayes

- Language : C#
- Output : Crop yield in tonnes
- Training : 80%
- Testing : 20%
- Precision : For Paddy - 96%

  For Ragi - 98%

Result Analysis

Dept Menu

- Home
- Paddy Dataset
- Paddy Yield(NB)
- Ragi Dataset
- Ragi Yield(NB)

Paddy Yield Result Analysis!!!

Constraint	Naive Bayes Algorithm
Accuracy	96%
Time (milli secs)	2153
Correctly Classified	96%
InCorrectly Classified	4%

Dept Menu

- Home
- Paddy Dataset
- Paddy Yield(NB)
- Ragi Dataset
- Ragi Yield(NB)

Ragi Yield Result Analysis!!!

Constraint	Naive Bayes Algorithm
Accuracy	98%
Time (milli secs)	503
Correctly Classified	98%
InCorrectly Classified	2%

Demonstration

- IDE : **Visual Studio**
- Programming Language : **C#**
- Back End : **MS SQL Server**

Future Scope

- Improvisation of the current system to predict yield for wide range of crops.
- Enhance the system to predict crop yield for different regions.
- Addition of an extended feature to give the chemical combination of the fertilizers that can produce better yield with the given land and weather conditions

Conclusion

- Farmers can benefit from yield prediction to make informed management and financial decisions.
- The presented machine learning approach can be useful as it is a real time application and provides satisfying results for the current dataset.
- Helps Policy makers to make timely import and export decisions to strengthen national food security.

References

- [1]. Sally Jo Cunningham and Geoffrey Holmes.Developing innovative applications in agriculture using data mining, , Year of Publication:1999
- [2]. D Ramesh, B Vishnu Vardhan .Data Mining Techniques and Applications to Agricultural Yield Data, Year of Publication: 2013[
- [3]. Monali Paul, Santosh K. Vishwakar- ma, Ashok Verma .,Analysis of Soil Behaviour and Prediction of Crop Yield using Data Mining Approach, Year of Publication: 2015
- [4]. Ami Mistry, Vinita Shah. Brief Survey of data mining Techniques Applied to applications of Agriculture, Year of Publication: 2016
- [5]. Yogesh Gandge, Sandhya, A Study on Various Data Mining Techniques for Crop Yield Prediction,Year of Publication: 2017
- [6]. Md. Tahmid Shakoor, Karishma Rahman, Sumaiya Nasrin Rayta, Amitabha Chakrabarty. Agricultural Production Output Prediction Using Supervised Machine Learning Techniques, , Year of Publication: 2017

- [7]. Umid Kumar Dey, Abdullah Hasan Masud, Mohammed Nazim Uddin, .Rice Yield Prediction Model Using Data Mining,Year of Publication: 2017
- [8]. Kuljit Kaur. Kanwalpreet Singh Attwal. Effect of Temperature and Rainfall on Paddy Yield using Data Mining, Year of Publication: 2017.
- [9]. Nidhi H Kulkarni, Dr. G N Srinivasan, Dr. B M Sagar, Dr.N K Cauvery. Improving Crop Productivity Through A Crop Recommendation System Using Ensembling Technique, Year of Publication: 2018.
- [10]. Waleed Khan, Nasru Minallah,Imran Ullah Khan, Zahid Wadud,Muhammad, Zeeshan,Suhail Yousaf And Abdul Baseer Qazi . On the Performance of Temporal Stacking and Vegetation Indices for Detection and Estimation of Tobacco Crop,Year of Publication: 2018
- [11]. Corn Belt, Author: Jie Sun , Zulong Lai, Liping Di , Senior Member, IEEE, Ziheng Sun, Jianbin Tao, and Yonglin Shen . Multilevel Deep Learning Network for County-Level Corn Yield Estimation in the U.S. Year of Publication: 2020

- [12] Neha Rale, Raxitkumar Solanki, Doina Bein, James Andro-Vasko, Wolfgang Bein. “ Prediction of Crop Cultivation” ,IEEE Jan 2019.
- [13]S.Bhanumathi, M.Vineeth and N.Rohit.”Crop Yield Prediction and Efficient use of Fertilizers”,IEEE April 2019.
- [14] Dhivya Elavarasan , P. M. Durairaj Vincent,“Crop Yield Prediction Using Deep Reinforcement Learning Model for Sustainable Agrarian Applications”,IEEE May 2020.
- [15] Dr. Y. Jeevan Nagendra Kumar, V. Spandana, V.S. Vaishnavi, K. Neha,”Supervised Machine learning Approach for Crop Yield Prediction in Agriculture Sector”,IEEE July 2020.

Thank you.