Crop Yield Prediction Using Machine Learning Techniques

[1]Prof. Satish J. Manje, [2]Mr. Shubham Samanta, [3]Mr. Ajay Gupta, [4]Mr. Manthan Bhave
[1]Asst.Professor,[2,3,4]UG Student,[1,2,3,4]Computer Engg. Dept. Shivajirao S. Jondhle College Engineering & Technology, Asangaon, Maharashtra, India. [1]satishmanje93@gmail.com, [2]shubhamsamanta11@gmail.com, [3]guptaajay28799@gmail.com, [4]manthanbhave99@gmail.com

Abstract- Agriculture is a major part of our country's economy. Agriculture gave birth to civilization. India is a crop-based country and its economy largely dependent upon agricultural productivity. Hence, agriculture is the foundation of all business in our country. Crop selection is beneficial for agriculture planning. The selection of crops will depends' upon multiple entities such as production rate, market price and the different government policies. Improving agricultural field will result in boosting countries economic status. Implementing machine learning technique, a drastic positive growth can be seen in farming sector. In order to get the accurate prediction precise information about the matters is also crucial. The basic concept is to resolve the farmers' agricultural problems. Thus, improvement in agriculture by maximizing crop yield production rate boost the Indian economy.

Keywords- Indian Agriculture, Machine Learning Techniques, Crop selection method.

I. INTRODUCTION

To achieve supreme yield rate of crops by using available land resource is the main goal in agricultural planning. Numerous machine learning algorithms can help to upgrade the crop production. Whenever there is a mislaying in unfavorable condition it can apply crop selecting method and minimize the losses. And it can achieve the crop yield rate in favorable condition. This results in increasing the yield and enhance the country's economy. It has some of the element which controls the crop yield rate. It needs test the quality of the seeds before sowing. As we know that good quality of seeds helps in getting more yield rate. The choices of crop dependency are favorable and unfavorable conditions. This can also be improved by using hybridization methods. Many, analysis carried out to raise agricultural planning. The goal is to get the maximal yield of crops. Machine learning techniques can be used to revamp the yield rate of crops. The method of crop selection is applied to improve crop production. The production of crops may depend on geographical circumstances of the region like river ground, hill areas or the depth areas. Weather entities like temperature, cloud, rainfall, humidity. Types of soil might be saline, peaty, sandy or clay. Soil comprises of potassium, nitrogen, phosphate, manganese, iron, copper, calcium. Many variables are used for multiple crops to do various predictions. These predicted result models can be analyzed by various researches.

II. AIMS AND OBJECTIVE

a) Aim

The nature and magnitude of agricultural sector depends on the evolution of the climate system, as well as the bridge between crop yields and weather. This paper aims on better yields prediction from climatic conditions. Accurate mapping weather to crop are not only for prominent impacts to agriculture, but also for prominent influence of climate variation on linked economic and environmental outcomes and in turn for mitigation and adaptation policy. Machine Learning (ML) technology has been improved with time. Fusion of machine algorithms with agricultural processes will help uplifting the farming potential.[1]

b) Objective

To obtain practicality, data training models considered in the proposed scheme. To be more specific, objectives of the proposed scheme are described as follows, to obtain estimates of aggregate physical production function for yields of various crops in specified states, considering various technological factors and newly developed better yield goals. It attempts to solve concerns by constructing up a model of an interactive prediction system to increase the accuracy of crop yield prediction. By analyzing all these concerns like temperature, humidity, rainfall, weather, moisture, there is no exact solutions and technologies to vanquish the circumstances faced by farmers.

III. LITERATURE SURVEY

The literature survey deals with the topics and the researches that would help to understand the existing systems that are similar to this project. The objective of this literature survey is to analyze the related work to this project and mechanisms used in previous studies.

Paper 1: CROP YIELD PREDICTION AND EFFICIRNT USE OF FERTILIZERS:

Analyze the various related attributes like location, pH value from which alkalinity of the soil is determined. Along with it, percentage of nutrients like Nitrogen (N), Phosphorous (P), and Potassium (K) Location is used along with the use of third- party applications like APIs for weather and temperature, type of soil, nutrient value of the soil in that region, amount of rainfall in the region, soil composition can be determined. All these attributes of data will be analyzed, train the data with various suitable machine learning algorithms for creating a model. The system comes with a model to be precise and accurate in predicting crop yield and deliver the end user with proper recommendations about required fertilizer ratio based on atmospheric and soil parameters of the land which enhance to increase the crop yield and increase farmer revenue.[2]

Paper2: SPIKING NEURAL NETWORKS FOR CROP YIELD ESTIMATION BASED ON SPATIOTRMPORAL ANALYSIS OF IMAGE TIME SERIES:

This paper illustrates this concept with the introduction of the first SNN computational model for crop yield estimation from normalized difference vegetation index image time series. It presents the development and testing of a methodological framework which utilizes the spatial accumulation of time series of Moderate Resolution Imaging Spectroradiometer 250-m resolution data and historical crop yield data to train an SNN to make timely prediction of crop yield. The research work also includes an analysis on the optimum number of features needed to optimize the results from our experimental data set.[3]

Paper 3: ESTIMATING CROP YIELDS WITH DEEP LEARNING AND REMOTELY SENSED DATA:

This paper describes Illinois corn yield estimation using deep learning and another machine learning, SVR. Deep learning is a technique that has been attracting attention in recent years of machine learning, it is possible to implement using the Caffe. High accuracy estimation of crop yield is very important from the viewpoint of food security. However, since every country prepare data un-homogeneously, the implementation of the crop model in all regions is difficult. Deep learning is possible to extract important features for estimating the object from the input data, so it can be expected to reduce dependency of input data.[4]

Paper 4: CROP YIELD PREDICTION WITH RANDOM FOREST DECISION TREES:

This paper focuses on predicting the yield of the crop by applying various machine learning techniques. The outcome of these techniques is compared on the basis of mean absolute error. The prediction made by machine learning algorithms will help the farmers to decide which crop to grow to get the maximum yield by considering factors like temperature, rainfall, area, etc. By combining rainfall, temperature along with other parameters like season and area, yield prediction for a certain district can be made. Results reveals that Random Forest is the best classifier when all parameters are combined. This will not only help farmers in choosing the right crop to grow in the next season but also bridge the gap between technology and the agriculture sector.[5]

IV. EXISTING SYSTEM

There are several methods of yield forecasting. The traditional method of yield forecasting is the evolution of crop status by experts. Observations and measurements are made with all over the growing crop season, such as tiller number, spikelet number and their fertility percentage, percentage of damage from pests and fungi, percentage of weeds infestation and so on. From the data obtained in the way yield can be forecasted using regression methods, or by the knowledge from local expertise. Other two methods used to forecast crop yield are the use of outlying observations and crop simulation models. The objective of yield forecast is to give a precise, scientific sound and independent forecast of crops' yield expeditiously possible during the crops' growing season by considering the effects of weather and climate. The differences between forecasts and final estimates are in the schedule of release. If farmers know that subsequent week there is a good chance for rain, then they will rush into the field to sown their seeds. To Forecast crop yield, includes more aspects.

V. COMPARTIVE STUDY

Sr. No	Paper Name	Author/ Publication	Technology	Advantages	Disadvantages
1.	Crop Yield prediction and efficient use of fertilizers.	S. Bhanumathi, M. Vineeth and N Rohit	The model comprises of KNN	Helps to obtain better yield prediction and crop production	This algorithm would require more data for processing and analyzing the results.

2.	Spiking Neural Networks for Crop Yield Estimation Based on Spatiotemporal Analysis of Image Time Series.	Pritam Bose, Nikola K. Kasabov, Lorenzo Bruzzone, and Reggio N. Hartono	The system uses SNN computational model for crop yield estimation from various normalized vegetation aspects	This study helps in successful yield estimation model using the proposed SNN technique.	The algorithm produces the output in some incomplete or in sufficient data which makes the inconsistent results.
3.	Estimating crop yields with deep learning and remotely sensed data.	Kentaro Kuwata and Ryosuke, Shibasaki	Illustrative data that uses support vector machine algorithm for estimation crop yield.	The results indicate that deep learning is possible to predict crop harvest yield.	The algorithm used in this project faces the problem called 'Blackbox'. Also, a huge data.
4.	Crop Yield Prediction with Random Forest Decision Trees.	Aruvansh Nigam, Saksham Garg, Archit Agrawal, Parul Agrawal	Predicted calculations through mean absolute error in random forest.	Helpful for the farmer make choice to grow next season, hence connecting agriculture with technology.	For every large dataset, the size of trees can take up lot of memory also faces a problem called Blackbox.

VI. PROBLEM STATEMENT

- In India farmers still follow the traditional technology which they adopted from their ancestor. But the problem is that in earliest time climate was very healthy everything was happened on time. But now most of things have been changed due to global warming and many other factors. The main problem with agriculture in India is lack of rainfall in seasonal time.
- Humidity is also necessary for crops but it has been excessive, it also converts as a drawback. Winter season has been affected so Rabi crops are widely affected. Since few years rainfall in winter season was high as expected.
- To overcome this problem, we need to develop a system which able to find the hidden facts or results, patterns and insights. The farmer can predict which crop should sow such that it can get more benefit.
- In Proposed system we are applying data analytic techniques on agricultural production-based datasets and find the insights such that it can help to the farmers and their decision making.
- Traditional method helps in predicting single sample spaces. And machine learning methods helps in predicting multiple predictions. We need not to consider the structure of data models in traditional method whereas we need to consider the structure of data models in machine learning methods.

VII. PROPOSED SYSTEM

Here the system going to use two different methods. First is Naive Bayes method and second is K-Nearest neighbor method. system can get the accuracy of performance by using these two methods. To speculate the crop yield rate an application is created. This application includes three parts. First is managing datasets second is testing datasets and third is analyzing the datasets. In managing datasets, system can get the datasets of previous years and they can also be converted into supporting format in testing part we can do the single testing. System considered two methods of machine learning. One is Naive Bayes and other is K-Nearest neighbor method. In testing user can select any one of the techniques and do testing of dataset like by selecting particular crop, particular place and particular season we can get results of yield. In analyzing part, user can input a whole dataset file and get accuracy of the two different methods. This helps in predicting which method is good. As, the farmers are facing many problems in agricultural sector, we need to minimize their problems. This reduction of problems can be done by implementing new techniques on agriculture. It was possible implement the machine learning methods on agriculture. The clustering and classification methods that can be applied on crops. Also, some of regression approach are used to improve the production of crops. Forecasting crop yield means also knowing of forecasting other important parameters. For example, quantifying the area of planted at the starting of growing season and quantifying the area harvested.

VIII. ALGORITHM

STEP 1: START

STEP 2: Get the input data set and requirements from user.

STEP 3: Process the dataset and convert into consistent form required.

STEP 4: Load and train the dataset (40% testing and 60% training).

STEP 5: Selected no of neighbors are initialized with K.

STEP 6: For every illustration in the data

STEP 6.1: Calculate the distance between the query illustration and the present illustration from the data.

STEP 6.2: Add the interval and index of instance to an ordered collection.

STEP 7: Sort the ordered collection of intervals and indices in increasing order (in ascending order) by the distances.

STEP 8: Select the first K inputs from the sorted collection.

STEP 9: Find the labels of the computed K inputs.

STEP 10: If regression, return the mean of the K labels.

STEP 11: If classification, return the mode of the K tags.

STEP 12: Calculate the results from the algorithmic models.

STEP 13: Save the predicted results and data for future analyses on local file base.

STEP 14: Display predicted yield results to the user.

STEP 15: algo naïve bayes

```
class Naïve Bayes:
    def __init..(self, name, crops):
        self.name <- name
        self.crops <- crops
known_yeilds <- use data sets form various states
```

STEP 16: Algo user crop analysis

```
def UserSendCropanalysis(request):
```

STEP 17: if crop in yeilds.crops:

```
possible.append(yeilds.name)
for x in possible:
        print('yeild is = ', x)
```

STEP 18: Algo yield details

```
def yeilddetails(request):
    dict <- cropyieldanalysismodel.objects.filter(email
<-email,status<-sts).order_by('-id')
```

STEP 19: X <- dataset

```
    y <- dataset1
```

STEP 20: X_ train, X_ test, y_ train, y_ test

```
<- train _test_ split(X, y, test _size<- 1 / 3,
random_ state<-0)
    st_X <- Standard Scaler()
    X_ train <- st _X .fit_ transform(X_ train)
    X_ test <- st_ X. transform(X_ test)
    classifier <- K Neighbors Classifier()
    classifier. fit(X_ train, y_ train)
    y_pred <- classifier.predict(X_test)
    print("y_pred", y_pred)
    cm <- confusion_matrix(y_test, y_pred)
    print("cm", cm)
    X_set, y_set <- X_train, y_train
    X1, X2 <-
nm.meshgrid(nm.arange(start<-X_set[:, 0].min() - 1,
stop<-X_set[:, 0].max() + 1, step<-0.01),
nm.arange(start<-X_set[:, 1].min() - 1, stop<-X_set[:,
1].max() + 1, step<-0.01))
    mtp.contourf(X1, X2,
classifier.predict(nm.array([X1.ravel(),
X2.ravel()]).T).reshape(X1.shape),
```

STEP 21: END

IX. MATHEMATICAL MODEL

A. Naive Bayes Algorithm

Naive bayes algorithm is completely based on naive bayes classifier. This classifier helps in finding the probability of predicted classes. This method is easy in building large datasets.

$$P(C|X) = \frac{P(X|C) * P(C)}{P(X)}$$

Bayes theorem allows to calculate succeeding probability _P(C|X) of the given P(X|C), _P(X) and _P(C).

P(C|X) = conditional probability of X when given C that is the succeeding probability.

P(X|C) = conditional probability of C when given X that is likelihood.

P(C) = prior probability of C.

P(X) = probability of X.

Naïve Bayas Classifier

$$P(a_i|V_i) = \frac{n_c + mp}{n + m}$$

- n is number of training examples for which v=vj.
- nc is number of examples for which v=vj and a=aj.
- p is prior estimation for P(ai |vj).
- m is equivalent sample size.

B. The Mathematics Behind KNN

As known, KNN works because of the deeply rooted mathematical theories it uses. When implementing KNN, the initial step is converting data points into attribute vectors, or their mathematical value. The algorithm then works by finding the range between the mathematical values of these points. The general approach to find this distance is the Euclidean distance, as shown below.

$$d(p,q) = d(q,p)$$

$$= \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \cdots + (q_n - p_n)^2}$$

$$= \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2}$$

KNN runs this formula to compute the interval between each data point and the test data. It calculates the probability of these points being related to the test data and classifies it based on which points share the highest probabilities. Breaking it down coded modules of KNN Get the most frequent class of these rows.

X. SYSTEM ARCHITECTURE

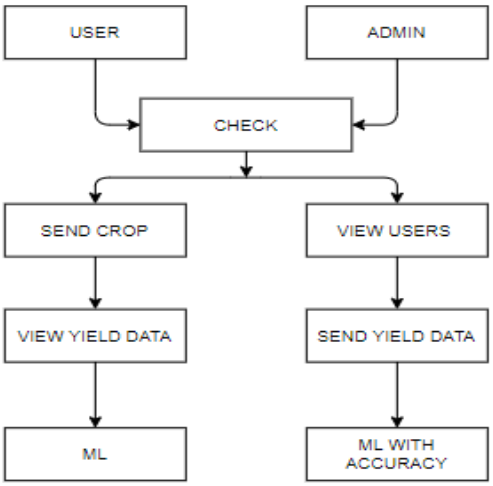

Fig.1: System Architecture

Description:

Here first we collect the data sets and process the data and we remove if there are any impurities in the data sets. Next the data is normalized if needed like it can be changed to smaller volume of data. Next the data is converted to supporting format. And then it is stored in the databases. Next the required method is applied. Now we get the final results. In this project all the datasets are converted to attribute relation file format. In testing part, we can do the single testing. We have considered two methods of machine learning. One is Naive bayes and other is K-Nearest neighbor method. In testing we can select any of process and do testing of datasets like by selecting particular crop, particular place and particular season we can get results of yield. In analyzing part, we can input a whole dataset file and get accuracy of the two different methods. This helps in predicting which method is good.

XI. ADVANTAGES

1. The system uses Machine Learning Algorithm and give best outcome based on precision. The result algorithm will be compared, best and accurate output will be selected.

2. This project suggests and implement rule-based system to forecast the crop yield production from the pool of past collected data.

3. It allows the real-time forecasting throughout the year and is applicable worldwide, especially for developing countries where field survey is hard to conduct.

4. The feature selection perspective successfully found important characteristics, and disclosed that environmental components had a greater effect on the crop yield than genotype.

5. It will act as a channel to provide the farmers a productive information required to get high yield and hence, increase profit which in turn will lessen the suicide rates and their difficulties.

6. The most practical issue in agriculture is making sure the crop yield production is maintained on an increasing scale by using algorithm to calculate the crop production and take necessary action if needed to ensure the yield will give higher profit.

XII. DESIGN DETAILS

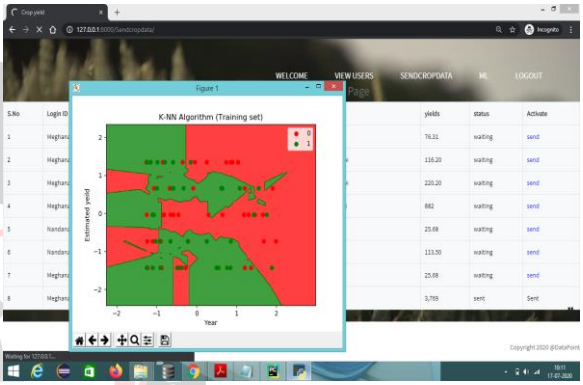

Fig 1: Graph on Crop Data

Fig 2: Final Accuracy Results

Analysis Results:

Analysis cases	**Crop details Name, Region, Season**	**Yield**
Case 1	Rice, Tamilnadu, Kharif	76.31
Case 2	Wheat, Uttarpradesh, Rabi	116.20
Case 3	Rice, Uttarpradesh, Kharif	220.20
Case 4	Wheat, Uttarakhand, Rabi	882
Case 5	Wheat, Haryana, Rabi	25.68

Case 6	Rice, Haryana, Kharif	113.50
Case 7	Rice, Maharashtra, Kharif	32.37
Case 8	Rice, Vellore, Kharif	3769

XIII. CONCLUSION

Thus, we have tried to implement "Crop Yield Prediction using Machine Learning Techniques." By Ramesh Meda, Vijay S. Rajpurohit, Shweta, IEEE 2019 5th International Conference for Convergence in Technology (I2CT) Pune, India. Mar 29-31, 2019. And the conclusion is as follows: These techniques also help in solving problems of agriculture. It gives the accuracy of yield by checking for different methods. Hence, it can improve the performance by checking the accuracy between different crops. This method can help in getting maximum yield rate of the crops. Also helps in selecting proper crop for their selected land and selected season.

REFERENCE

[1]Ramesh Meda, Vijay S. Rajpurohit, Shweta. "Crop Yield Prediction using Machine Learning Techniques.", IEEE 2019 5th International Conference for Convergence in Technology (I2CT) Pune, India. Mar 29-31, 2019.

[2]S. Bhanumathi, M. Vineeth and N. Rohit "Crop Yield Prediction and Efficient use of Fertilizers", IEEE 2019. International Conference on Communication and Signal Processing, April 4-6, 2019, India.

[3]Pritam Bose, Nikola K. Kasabov, Lorenzo Bruzzone, and Reggio N. Hartono "Spiking Neural Networks for Crop Yield Estimation Based on Spatiotemporal Analysis of Image Time Series", IEEE 2016. IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.

[4]Kentaro Kuwata and Ryosuke Shibasaki "Estimating crop yields with deep learning and remotely sensed data", IEEE 2015. The University of Tokyo IIS, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, JAPAN

[5]Aruvansh Nigam, Saksham Garg, Archit Agrawal, Parul Agrawal "Crop Yield Prediction with Random Forest Decision Trees", IEEE 2019. 2019 Fifth International Conference on Image Information Processing (ICIIP).

