Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 2551-2557
Received 25 April 2021; Accepted 08 May 2021.

Crop Yield Prediction by using Machine Learning Techniques

K.Gangadhara Rao[1],Assistant Professor, Keerthi Yashwanth[2] , M.Sridhar Goud[3]

Department of Information Technology,Chaitanya Bharathi Institute of Technology,Hyderabad,Telangana,India.

Abstract: Agriculture plays a significant role in the growth of the national economy. It relay on weather and other environmental aspects. Some of the factors on which agriculture is dependent are Soil, climate, flooding, fertilizers, temperature, precipitation, crops, insecticides and herb. The crop yield is dependent on these factors and hence difficult to predict. To know the status of crop production, in this work we perform descriptive study on agricultural data using various machine learning techniques. Crop yield estimates include estimating crop yields from available historical data such as precipitation data, soil data, and historic crop yields. This prediction will help farmers to predict crop yield before farming. Here we are utilizing three datasets like as clay data set, precipitation data set, and production data set of Telangana state, then we structure an assembled data sets and on this data set we employ three different algorithms to get the genuine assessed yield and the precision of three different methods. K-Nearest Neighbor(KNN) are applied on the training data set and are tested with the test data set, and the implementation of these algorithms is done using python programming and spyder tool.The proposed model can exhibit the precise expense of assessed crop yield and itis mark like as LOW, MID, and HIGH.

Keywords: Agriculture, Crop yield,Crop prediction,Machine learning,K-Nearest Neighbour.

I INTRODUCTION

The main goal of agricultural planning is to achieve maximum yield rate of crops by using limited number of land resources. Many machine learning algorithms can help in improving the production of crop yield rate. Whenever there is loss in unfavourable conditions we can apply crop selecting method and reduce the losses. And it can be used to gain crop yield rate in favourable conditions. This maximizing of yield rate helps in improving countries economy. We have some of the factors that influence the crop yield rate. They are seed quality and crop selection. We need test the quality of the seeds before sowing. As we know that good quality of seeds helps in getting more yield rate. And selection of crops depends upon two things that is favourable and unfavourable conditions. This can also be improved by using hybridization methods.

Many researches are carried out to improve agricultural planning. The goal is to get the

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 2551-2557
Received 25 April 2021; Accepted 08 May 2021.

maximum yield of crops. Many classification methods are also applied to get maximum yield of crops. Machine learning techniques can be used to improve the yield rate of crops. The method of crop selection is applied to improve crop production. The production of crops may depend on geographical conditions of the region like river ground, hill areas or the depth areas. Weather conditions like humidity, rainfall, temperature, cloud.Soil type may be clay, sandy, saline or peaty. Soil composition can be copper, potassium, phosphate, nitrogen, manganese, iron, calcium, ph value or carbon and different methods of harvesting. Many parameters are used for different crops to do different predictions. These prediction models can be studied by using researches. These predictions are classified as two types. One is traditional statistic method and other is machine learning techniques. 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 where as we need to consider the structure of data models in machine learning methods.

1. Geography of region,
2. Weather (Temprature, humidity ,percipitation),
3. Soil type (saline, alkaline,sodic,non-alkaline),
4. Soilcomposition(ph,N,P,K,EC,OC,Zn,F).

II LITERATURE REVIEW

Machine learning in Agriculture is a Novel field, a great deal of work has been done in field of Agriculture utilizing Machine learning. There are diverse guaging philosophies created and assessed by the specialists everywhere throughout the world in the field of farming or related sciences. Agricultural scientists in Pakistan have demonstrated that endeavors of harvest yield amplification through expert pesticide state strategies have prompted a hazardously high pesticide use. These examinations have revealed negative relationship between's pesticide use and harvest yield [5]. In their investigation they have explained that how data mining incorporated farming information including irritation exploring, pesticide utilization and meteorologicalinformation are helpful for streamlining of pesticide use.

Topical data identified with agribusiness which has spatial properties was accounted for in one of the study[6]. Their research went for perceiving patterns in farming creation with references to the accessibility of information assets. K-means method turned into applied to carry out gauges of the contamination in the air[7],the k- nearest neighbor become connected for mimicking day by day precipitations and other climate elements [8],and numerous ability changes of the weather situations are dissected utilizing SVM[9]. Statistics mining techniques are often used to have a look at soil qualities. As example, the k- means method is used for segmenting soils in mixture with GPS-based technology [10]. A decision

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 2551-2557
Received 25 April 2021; Accepted 08 May 2021.

tree classifier for agriculture information turned into proposed [11].This new classifier uses new facts expression and can address each entire records and in entire records. Inside the test,10-fold cross validation technique is used to check the data set, horse-colic data set and soybean data set. Their results showed the proposed selection tree is capable of classifying all styles of agriculture records. A yield prediction version turned into proposed in one of the take a look at [12] which makes use of data mining techniques for category and prediction. This model worked on enter parameters crop name, land location, soil type, soil ph, pest information, climate, water stage, seed type and this model anticipated the plant boom and plant diseases and therefore enabled to select the nice crop based on climate information and required parameters.

There are few research works about sugarcane yield prediction which can be associated with our work. Sugarcane yield prediction technique with use of Random forest [13] became proposed in one of the survey, the features used in this study consist of biomass index, climate statistics (e.g., rainfall) and yields from previous years. Two predictive tasks are provided in [13]: (i) the category problem for predicting whether or not the yield can be above or underneath the found median yield, and (ii) the regression hassle for predicting the yield estimates in two distinct time intervals. In addition, support vector system[14] for rice crop yield prediction become proposed, the data set used in this method are precipitation ,minimum, maximum and common temperature, place and manufacturing. The sequential minimal optimization classifier is implemented on the data set.

III PROPOSED SYSTEM/METHODOLOGY

In the proposed system, we use supervised learning to form a model, which provides predicted cost of crop yield and corresponding production order. The proposed system is described in following stages such as dataset collection, pre-processing step, feature selection and applying machine learning modules as shown in figure 1.

Data set Collection*:* Data is collected from a variety of sources and prepared for data sets. And this data is used for descriptive analysis. Data is available from several online abstract sources such as Kaggle.com and data.gov.in. We will use an annual summary of crops for at least 10 years. The data sets used in this paper are soil data set, rainfall data set and crop yield data.

Pre processing step*:* This step is a very important step in machine learning. Preprocessing consists of inserting the missing values, the appropriate data range, and extracting the functionality. The kind of the dataset is critical to the analysis process. In this paper we have used isnull() method for checking null values and lable Encoder() for converting the categorical data into numerical data.

Feature Selection: Feature extraction should simplify the amount of data involved to represent a large data set. The soil and crop characteristics extracted from the pre-treatment phase

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 2551-2557
Received 25 April 2021; Accepted 08 May 2021.

constitute the final set of training. These characteristics include the physical and chemical properties of the soil. Here, we have used Random Forest Classifier() method for feature selection. This method selects the features based on the entropy value i.e., the attribute which is having more entropy.

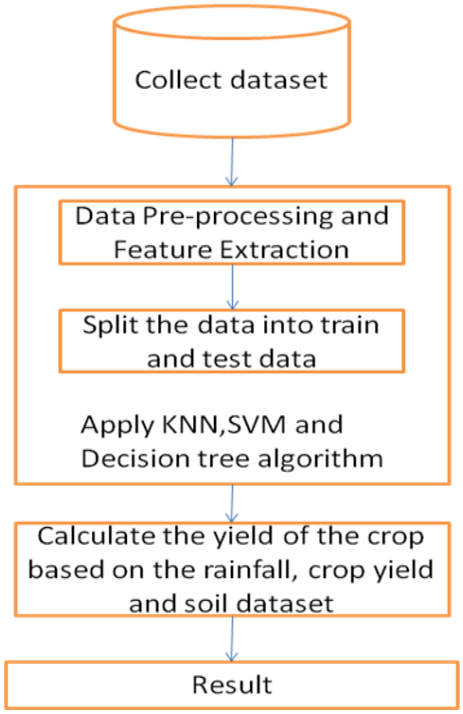

Fig.1 System Architecture

Split the Dataset into Train and Test Set*:* This step includes training and testing of input data. The loaded data is divided into two sets, such as training data and test data, with a division ratio of 80% or 20%, such as 0.8 or 0.2. In a learning set, a classifier is used to form the available input data. In this step, create the classifier's support data and preconceptions to approximate and classify the function. During the test phase, the data is tested.The final data is formed during preprocessing andis processed by the machine learning module.

Applying Machine Learning modules**:** In our project we have used three different supervised machine learningalgorithms for crop yield prediction which is given as follows

KNN Algorithm

KNN is a nonparametric supervised learning technique that uses training sets to segment data points into given categories. In simple classifications, the word collects information from all educational cases and similarities based on the new case. Look at the training for the most similar (neighbor) K cases and predict the new instance (x) by summarizing the output variables for these K cases. Classification is the class value mode (or most commonly). A flow diagram of the KNN algorithm is shown in Figure 2.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 2551-2557
Received 25 April 2021; Accepted 08 May 2021.

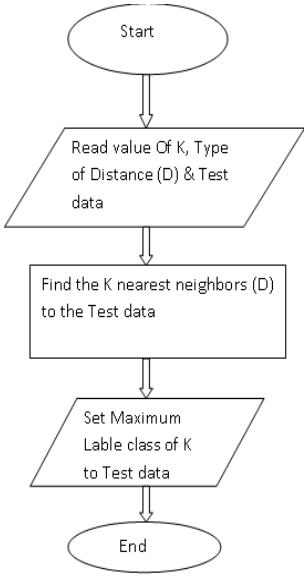

Fig.2 Flow chart for KNN algorithm

IV. RESULTS AND DISCUSSION

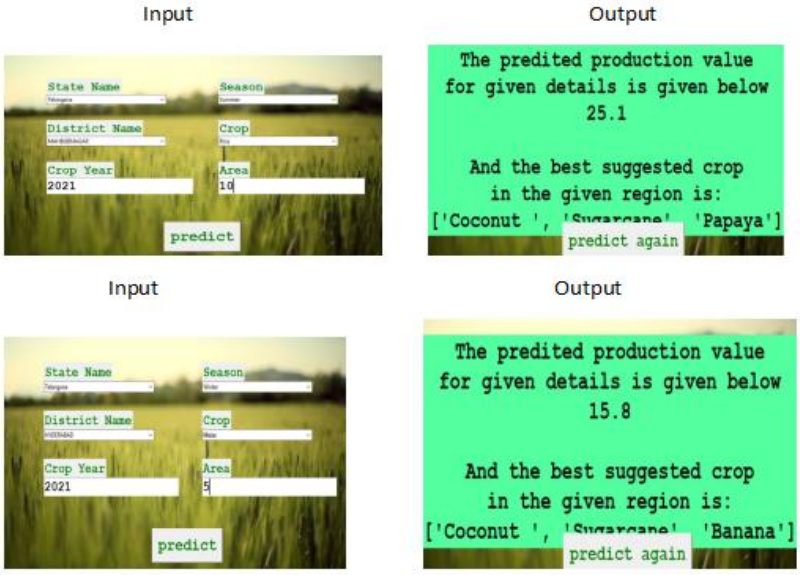

Fig. 3 . Home page

Fig. 3 shows the home page of the website where the person accessing the website enters the details such as the district, crop, season and the area in Hectare and by clicking on predict the result is printed

.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 2551-2557
Received 25 April 2021; Accepted 08 May 2021.

District Name	Season	Crop	PJan	PFeb	PMar	PApril
Mahabubnagar	Kharif	Tur	3.099	0	1.67	25.1
Mahabubnagar	Kharif	Bajra	3.099	0	1.67	25.1
Mahabubnagar	Kharif	Gram	3.099	0	1.67	25.1
Mahabubnagar	Kharif	Jowar	3.099	0	1.67	25.1
Mahabubnagar	Kharif	Maize	3.099	0	1.67	25.1
Mahabubnagar	Kharif	moong	3.099	0	1.67	25.1
Mahabubnagar	Kharif	pulses	3.099	0	1.67	25.1
Mahabubnagar	Kharif	Ragi	3.099	0	1.67	25.1
Mahabubnagar	Kharif	Rice	3.099	0	1.67	25.1
Mahabubnagar	Kharif	sugar cane	3.099	0	1.67	25.1
Mahabubnagar	Kharif	Total food grain	3.099	0	1.67	25.1
Mahabubnagar	Kharif	Urad	3.099	0	1.67	25.1
Mahabubnagar	Rabi	jowar	3.099	0	1.67	25.1
Mahabubnagar	Rabi	Maize	3.099	0	1.67	25.1
Mahabubnagar	Rabi	Other pulses	3.099	0	1.67	25.1
Mahabubnagar	Rabi	Wheat	3.099	0	1.67	25.1
Mahabubnagar	Summer	Maize	3.099	0	1.67	25.1

Fig. 4. Data set

Fig. 4. It is the snapshot of the final processed data set that is being used for this project

CONCLUSION AND FUTURE SCOPE

Different machine learning algorithms have been implemented on agricultural data to evaluate the best performing method..Based on the climatic input parameters the present study provided the demonstration of the potential use of data mining techniques in predicting the crop yield based. The developed web page is user friendly and the accuracy of predictions are above 75 per cent in all the crops and districts selected in the study indicating higher accuracy of prediction. By providing climatic data of that place the user-friendly web page developed for predicting crop yield can be used by any user their choice of crop.

REFERENCES

1. Arun Kumar, Naveen Kumar, Vishal Vats.”Efficient crop yield prediction using machine learning algorithms”, IJRET Volume: 05 Issue: 06, June-2018, pp 3151-3159
2. P.Priya, U.Muthaiah ,M.Balamurugan**. “** Predicting yield of the crop using machine learning Algorithm”, IJESRT *et al.,* 7(4)**:** April-2018, pp 2277-2284
3. Vaneesbeer Singh, Abid Sarwar, Vinod Sharma. **“**Analysis of soil and prediction of crop yield (Rice) using Machine Learning approach”, IJARCS 8 (5), May-June 2017, pp 1254-1259.
4. A.S. Ponraj, Vigneswaran T. “Machine Learning Approach for Agricultural IoT”,IJRTE, Volume-7, Issue-6, March 2019,pp 383-391.
5. Abdullah, A., Brobst, S, Pervaiz.I., Umer M.,and A.Nisar, “Learning dynamics of pesticide abuse through data mining”, Proceedings of Australian Workshop on Data

mining and Web Intelligence, New Zealand, January .2004,pp 63-68.

6. Kiran Mai,C., Murali Krishna, I.V, an A.VenugopalReddy, " Data Mining o f Geospatial Database for Agriculture Related Application", Proceedings of Map India,New Delhi, 2006,pp 83-96.
7. Jorquera H, Perez R, Cipriano A, Acuna, "Short term forecasting of air pollution episodes",In. Zannetti P (eds) Environmental Modeling 4. G(2001) WITPress, UK,pp 221-237.
8. Rajagopalan B. Lall, " A k- nearestneighbor daily precipitation and other weather variables." U (1999) WatResResearch35(10) :3089 – 3101.
9. Tripathi S, Srinivas VV, Nanjudiah , "Down scaling of precipitation for climate change scenarios: a support vector machine approach", RS (2006),J. Hydrology 330-337.
10. Verheyen, K., Adrianens, M. Hermy and S.Deckers(2001). High resolution continuous soil classification using morphological soil profile descriptions. Geoderma, 101:31-48.