Expanding the Crop Yield with Machine Learning and IOT

Mr. Nilesh B. Patel[1], Dr. Mahammad Idrish I. Sandhi[2]
[1]*Assistant Professor, Shree Ambaji BCA College, Ambaji , Hem. North Gujarat University, Patan, Gujarat, India*
[2]*Dean & Hod, Faculty of Computer Science, Sankalchand Patel University, Visnagar, Gujarat, India*

Abstract— Agriculture is a key lucrative source for a healthier live hood. The demand for agriculture is always a boom in all Agriculture is aspects of our living. Country farmers should track the climate change, and cope up the demand for high and organic kind of food. In order to shoot up the growth and yield of crops production, the farmer should be knowledgeable on the seasonable climatic fluctuations, this factor will assist to conclude the choice of crop growth. IoT merged Smart Farming will definitely boost the system in Agriculture, where the field is monitored in real-time and gives a precise extraction and analysis of data. Machine learning has an immense power in crop production sector by raising the productivity and the quality of production. Selecting the relevant algorithm for the defined data will support harvesting an profitable crop.

Index Terms— Agriculture, IOT, Machine Learning, Crop

I. INTRODUCTION

Considered to be the primary occupation of our nation. Production is the root vein. Comparatively in recent years productivity of crops is moderate. As a result, the demand of food is increasing. Researchers are trying to put extra effort for more production. In past farmers used to follow naked eye observations. This method of naked eye examining is difficult as far as when large hectares of productions are considered. Hence in order to obtain finest results in predicting the crop yield, some of the machine learning algorithms can be implemented. It's the study of concepts which empower computers to do the things that make people seem intelligent. It is the learning through knowledge, experience regarding some tasks and performance

Food insecurity may be a drawback that can't be avoided. We humans are responsible for betterment in the method of cultivation considering its water, air and soil condition. There is a large gap of information sharing from ancient farmers to the new agricultural technologies, chances to overcome if the computer code may be designed to model the interactive impact of climate factors, particularly the impact of maximum events like temperature, rainfall and water level, occurring at completely different growing phases of crops. The temperature changes undoubtedly damage the native and overall food production, therefore planning computer code to design crop predictions needs latest methodology for temperature change studies, situations to adapt the temperature variations, and policymaker.

The soil sort will modify over time because of climatic condition and pests, therefore crop management must manage a fancy quantity of information, directly or indirectly associated with one another. It will therefore by considering a simplified reality, to permit a quick assessment of the impact of temperature change in agriculture.

Machine learning techniques focus extremely on characteristic correlations or patterns among massive relative databases. This is shared as a knowledge reborn into information which gets noted for futuristic knowledge gain. This information will be a backbone for farmers for the prediction of crop cultivation by predicting probabilities of crop losses or stop losses. Planning, learning, communication, reasoning, knowledge and perception are some of the central principles which is the ability to move and manipulate objects. It is the science and engineering of compiling intelligent machines. Modern technological advancement in the field of yield prediction may also aid farmers in cost prediction based upon the production. Mainly there are two categorizations in yield prediction, classification and prediction phase. Predictions can be obtained by examining large sets of pre- existent databases in order to generate new knowledge. It is also considered as one of the proven methods of resolving issues. Thus, resulting in an automatic and approximate predictions. Various steps that can be involved in crop yield prediction are stated below.

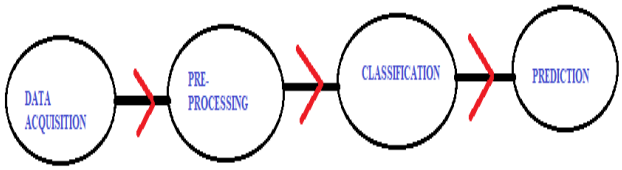

FIG.1 CROP YIELD PREDICTION

A. Data Acquisition

It is a process of collecting data of yield production and giving it as input. It needs two things to work, data and models. The acquired data should have enough features (aspect of data that can aid for a prediction).

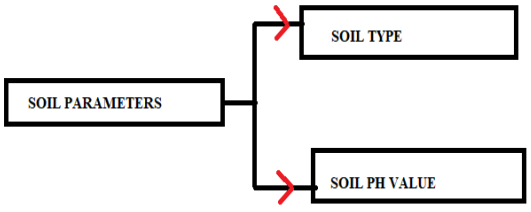

FIG.2 TYPES OF SOIL PARAMETER

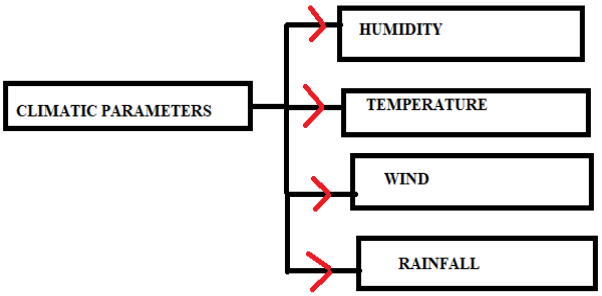

FIG 3. TYPES OF CLIMATIC PARAMETER

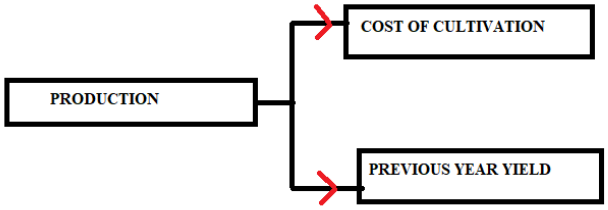

FIG 4. TYPES OF PRODUCTION

B. Pre-Processing Of Data

In Data pre-processing method the original datas are taken and filtered to define a error free data. In short, whenever the data is clustered from various sources, it is obtained in raw format which is not probable for the analysis.

The dataset that is used needs to be pre-processed because of the presence of redundant attributes, noisy data in it. Initially, data cleaning operation is performed where the redundant factors are determined and are not considered for the prediction of crops.

C. Classification

It is a process of determining, in which a set of categories the collected data has to be classified. From this new consideration data associated with class are defined.

D. Prediction

It is the process of referring to an algorithm's output, after it has been skilled on a factual dataset and applied to advanced data when forecasting the tendency of a particular outcome.

II. LITERATURE REVIEW

Nowadays many experts are applying automated farming. Since Decision Tree is a well-known algorithm, which predicts the model in generalized pattern. They made a trail on soybean, how the decision tree influences the climatic condition. Based on productivity. For easy understanding of end-user different kind of rules were created from the Decision tree. The paper from Md. Tahmid Shakoor & co paper helped us for selecting various attributes like land capability classification, soil depth, slope, drainage, and permeability [4].

Two classic method algorithms on machine learning have been introduced. In the invented system, required soil and climatic condition has being automatically organized by the given sources. Additionally, the system will work on huge landscape, and provides forecasts at a resolution with good given data resolution, similarly to the instance is initially from the soil data. Pre-forecasting the crop before the inception of the crop season will help the users with the capability to enhance changes, like deciding a high robust genetic difference before planting, the crop type, in order to put up for extreme climatic variations further ahead in the crop cycle [2].

The given algorithm developed a data-driven model that predicts and forecast the crop yield depending on soil and climate features. There are many algorithm has been suggested for rainfall prediction but, the algorithm discussed in this paper succeeded in accentuate on Rainfall along with the crop yield prediction. The designed algorithm mainly depends on environment and soil parameter for crop cultivation, which are taken into account for final stage of prediction. This will highly benefit the farmers by knowing the needed crop to be sown, before the sowing period.

This predictive pattern algorithm will exclusively help the local self-government and financial institutions to give considerable funds or fiscal loans to farmers. Use of naïve Bayes and decision tree makes the model very efficient especially for computation. It is scalable and can be used to trial on different crops. From the yield graphs, the best time of sowing, plant growth and harvesting of the plant can be found out.

III. EXISTING SYSTEM

The growth of our country totally depends on agriculture wealth. When a population of the country increases, indirectly it increases the necessity of agriculture which will subsequently affects the economic growth. At this instance, the crop yield rate plays a significant role in the profit-making growth of the country. To increase the rate of crop yield, some biological approaches and some chemical approaches have been raised to solve this issue. Additionally, crop sequencing technique has been slotted to improvise the net yield of the crop. We have taken example of CSM to demonstrate how it helps farmers in achieving more yield Crop can be classified as:

TABLE I TYPES OF CROPS

Types	Description	Example
Seasonal crop	Planted based on season.	Wheat, cotton
Whole year crop	Planted in entire season.	Paddy
Short time crop	Growth time is minimum.	Potato
Long time crop	Growth time is long.	Sugarcane

Basically, in crop selection method makes use of technique where it recommends different set of crops for same area over the years. There are various options are available to select for farmers. They can choose one of the options and observe the results. The combination which will give high yield for same area is generated as output for that area. In this way CSM is targeted to define the best crop for given land area. Farming Systems especially in India are strategically planned, in accordance with the locations where they are most gain able. The agricultural systems on farming are subsequently classified as:

- Subsistence farming,
- Organic farming,
- Industrial farming.

In India, the type of framing differs from region to region, some are based on horticulture, agro forestry. The surveyed research papers have given a rough idea about using ML with only one attribute. We have the aim of adding more attributes to our system and ameliorate the results, which can improve the yields and we can recognize several patterns for predictions. This system will be useful to justify which crop can be grown in a particular region.

IV. CLASSIFICATION OF MACHINE LEARNING

Modern technological advancements in the field of yield forecast may also aid farmers in cost prediction based upon the production. Mainly there are two varieties in yield prediction.,

classification and prediction phase. In this paper we are presenting an outline on classification and predictions techniques. Predictions can be acquired by analyzing the large sets of pre-existent databases in order to generate a new understanding. It is also considered as one of the proven methods of settling the issues. Thus, resulting in automatic and rough predictions. The different types of machine learning are:

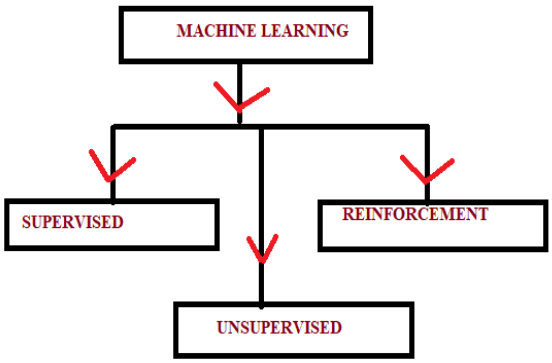

FIG 5. MACHINE LEARNING

SUPERVISED LEARNING

It is the machine learning task of provoking a function from an imposed training data. Needs the training method to deduce from the training resource to hidden situations during a sensible way.

The working principle of supervised model can be stated as, after the completion of learning, new data is generated without outputs, but the system get the capability to generate the output with the help of knowledge gained during the training segment. Supervised learning is often divided into

- Classification
- Regression Problems.

A. Classification

It predicts the outcome of a given sample and segregate the output variable based on categories.

B. Regression

The output is predicted from the given sample, where the defined output variables are in real numerical values. Some of the supervised machine learning algorithms are as follows:

C. Decision Trees

It is a specified for classification problems. More specifically, it is applicable for both categorical and continuous dependent variables.

D. Random Forest

Random Forest can be considered as an ensemble of decision trees. To categories a replacement object supported on attributes, every single tree will specify grouping and that we say the tree "votes" for that class. The forest will always chooses the group that always having the foremost votes (over all the trees within the forest).

Support Vector machine

In this work the supervised machine learning algorithm SVM is used for categorization of microaneurysms. Support Vector Machine might act as a supervised machine learning algorithm which supports both classification or regression and the challenges caused by it.

k - NN Classifier

k - NN classifier is being framed from lazy learning concept. In the early stage, classification is not being patterned. After the users input their prediction queries, generalization occur, based on distance function, the values are predicted.

UNSUPERVISED LEARNING

Unsupervised learning is where we simply have an input file (A) and it doesn't hold an corresponding output variables. It got classified as follows:

- Clustering
- Association Problems

A. Clustering

In clustering, the problem arises, when the inborn separation within the data, which resembles like grouping the customers based on their purchasing behavior.

B. Association

An association rule learning, here describes that rule are discovered which describes, major part of our data, were farmer purchase A and B products. Some verified examples of unsupervised learning algorithms are given below:

- k-means
- Apriori algorithm

REINFORCEMENTLEARNNING

This algorithm will permit the machines and software agents to automatically to define the characteristics within a noted context, so as to increase its enforcement.

V. METHODOLOGY

The proposed methodology contains of data acquisition, preprocessing of data, classification and prediction.

A. Data Acquisition

It is a process of collecting data of yield production and giving it as input. It needs two things to work, data and models. The acquired data should have enough features (aspect of data that can aid for a prediction). The parameters selected are year, month, temperature, humidity, rainfall, production, target. The rainfall directly affects the crop production. Here is the total amount of precipitation of crops of each month of every year. Year Wise temperature details are taken into consideration as one of the inputs. SHEET 1 DATA ACQUISITION of collecting data of yield production and giving it as input. It needs two things to work, data and models. The acquired data should have enough features (aspect of data that can aid for a prediction).

	A	B	C	D	E	F	G
1	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
2	Andaman and Nicob	NICOBARS	2000	Kharif	Arecanut	1254	2000
3	Andaman and Nicob	NICOBARS	2000	Kharif	Other Kharif pulses	2	1
4	Andaman and Nicob	NICOBARS	2000	Kharif	Rice	102	321
5	Andaman and Nicob	NICOBARS	2000	Whole Year	Banana	176	641
6	Andaman and Nicob	NICOBARS	2000	Whole Year	Cashewnut	720	165
7	Andaman and Nicob	NICOBARS	2000	Whole Year	Coconut	18168	65100000
8	Andaman and Nicob	NICOBARS	2000	Whole Year	Dry ginger	36	100
9	Andaman and Nicob	NICOBARS	2000	Whole Year	Sugarcane	1	2
10	Andaman and Nicob	NICOBARS	2000	Whole Year	Sweet potato	5	15
11	Andaman and Nicob	NICOBARS	2000	Whole Year	Tapioca	40	169
12	Andaman and Nicob	NICOBARS	2001	Kharif	Arecanut	1254	2061
13	Andaman and Nicob	NICOBARS	2001	Kharif	Other Kharif pulses	2	1
14	Andaman and Nicob	NICOBARS	2001	Kharif	Rice	83	300
15	Andaman and Nicob	NICOBARS	2001	Whole Year	Cashewnut	719	192
16	Andaman and Nicob	NICOBARS	2001	Whole Year	Coconut	18190	64430000
17	Andaman and Nicob	NICOBARS	2001	Whole Year	Dry ginger	46	100
18	Andaman and Nicob	NICOBARS	2001	Whole Year	Sugarcane	1	1
19	Andaman and Nicob	NICOBARS	2001	Whole Year	Sweet potato	11	33
20	Andaman and Nicob	NICOBARS	2002	Kharif	Rice	189.2	510.84

SHEET 1

20970	Assam	KARBI ANGLONG	2000	Kharif	Mesta	74	386
20971	Assam	KARBI ANGLONG	2000	Kharif	Niger seed	42	22
20972	Assam	KARBI ANGLONG	2000	Kharif	Sesamum	2426	1749
20973	Assam	KARBI ANGLONG	2000	Kharif	Small millets	206	66
20974	Assam	KARBI ANGLONG	2000	Rabi	Gram	547	272
20975	Assam	KARBI ANGLONG	2000	Rabi	Linseed	91	39
20976	Assam	KARBI ANGLONG	2000	Rabi	Other Rabi pulses	2875	1547
20977	Assam	KARBI ANGLONG	2000	Rabi	Rapeseed &Mustard	17138	8932
20978	Assam	KARBI ANGLONG	2000	Rabi	Wheat	1364	2146
20979	Assam	KARBI ANGLONG	2000	Summer	Rice	1559	1713
20980	Assam	KARBI ANGLONG	2000	Whole Year	Arecanut	1198	739
20981	Assam	KARBI ANGLONG	2000	Whole Year	Banana	1700	22817
20982	Assam	KARBI ANGLONG	2000	Whole Year	Coconut	391	1440000
20983	Assam	KARBI ANGLONG	2000	Whole Year	Dry chillies	291	162
20984	Assam	KARBI ANGLONG	2000	Whole Year	Onion	221	447
20985	Assam	KARBI ANGLONG	2000	Whole Year	Potato	824	6623
20986	Assam	KARBI ANGLONG	2000	Whole Year	Sugarcane	4267	175995
20987	Assam	KARBI ANGLONG	2000	Whole Year	Sweet potato	246	775
20988	Assam	KARBI ANGLONG	2000	Whole Year	Tapioca	293	1546

SHEET 2

49500	Chhattisgarh	DURG	2011	Rabi	Urad	153	38
49501	Chhattisgarh	DURG	2011	Rabi	Wheat	6755	7399
49502	Chhattisgarh	DURG	2011	Whole Year	Banana	89	2716
49503	Chhattisgarh	DURG	2011	Whole Year	Coriander	71	20
49504	Chhattisgarh	DURG	2011	Whole Year	Dry chillies	14	15
49505	Chhattisgarh	DURG	2011	Whole Year	Garlic	108	316
49506	Chhattisgarh	DURG	2011	Whole Year	Onion	42	160
49507	Chhattisgarh	DURG	2011	Whole Year	Sugarcane	98	1174
49508	Chhattisgarh	DURG	2011	Whole Year	Sweet potato	54	334
49509	Chhattisgarh	DURG	2011	Whole Year	Turmeric	6	7
49510	Chhattisgarh	DURG	2012	Kharif	Arhar/Tur	1147	655
49511	Chhattisgarh	DURG	2012	Kharif	Cotton(lint)	76	114
49512	Chhattisgarh	DURG	2012	Kharif	Horse-gram	4	1
49513	Chhattisgarh	DURG	2012	Kharif	Jowar	8	5
49514	Chhattisgarh	DURG	2012	Kharif	Maize	120	186
49515	Chhattisgarh	DURG	2012	Kharif	Moong(Green Gram)	37	10
49516	Chhattisgarh	DURG	2012	Kharif	Other Kharif pulses	3	1
49517	Chhattisgarh	DURG	2012	Kharif	Potato	1	6
49518	Chhattisgarh	DURG	2012	Kharif	Rice	129784	296124
49519	Chhattisgarh	DURG	2012	Kharif	Sesamum	216	56

SHEET 3

B. Pre-Processing Of Data

In data pre-processing raw data are filtered into a clean data set (i.e., reasonable format). In short, whenever the data is clustered from various sources, it is obtained in raw format which is not probable for the analysis. The aim of pre-processing is to improve the generalizability of the model. It handles many irrelevant and missing data. In this, pre-processing out by the in-built functions in MAT LAB. In this technique raw data are being modified into a clean data set.

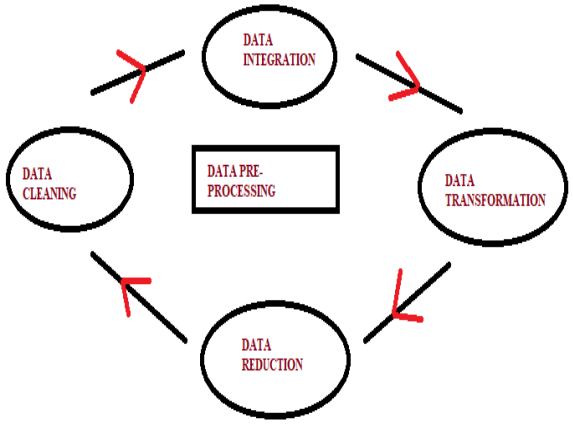

FIG 6. DATA PRE-PROCESSING

C. Classification

It has a flowchart-kind of structure where the internal node will represent a "test" on assigning, every single branch signals the result of the test, and each sub node signifies a class label. The diagrammatic representation of the tree structure gives us the best expected value to be calculated and it acts as a decision supportive tool. Following are the nodes of Decision tree,

- Decision nodes – squares representation
- Chance nodes – circles representation
- End nodes – triangles representation

It as a descriptive means for calculating conditional. In this method, data are given as input for the classification purpose, which gets processed by MATLAB and finally decision tree is being generated. To reduce complexity,

Pruning has been chosen. when the range of pruning starts increasing the complexity is reduced, where the performance level starts increasing.

D. Prediction

It is the process of referring to an algorithm's output, after it has been skilled on a factual dataset and applied to advanced data when forecasting the tendency of a particular outcome.

i. Prediction Using Artificial Neural Network

This Network is targeted to predict the software defect. A neural network has several concealed inputs, output nodes. Each node with match with similarity of some data, and turn backs an output. Till obtaining the output, the processing layer will consider the total average of the output of the very last layer. average of the outputs of the previous layer. The reason behind this concept is, multiple number of nodes will be able to solve the problem very easily that a single node operation finds quite impossible.

ii. Predicting the Crop Yield

Various regression algorithm has a remarkable effect on estimation on crop yield. Linear Regression and cross-validation model have been rejected because of negative scores. The Decision Tree Regressor is an algorithm achieved a good r2 score, when it gets used with the considered dataset, compared to other similar regression algorithms.

TABLE II
CROP PREDICTION ALGORITHMS

Algorithm	Accuracy on crop prediction
Decision tree classifier	99.87%
K-NN Classifier	99.73%
Support vector machine	92.6%
Random forest classifier	81.07%

The submitting author is responsible for obtaining agreement of all coauthors and any consent required from sponsors before submitting a paper. It is the obligation of the authors to cite relevant prior work.

Authors of rejected papers may revise and resubmit them to the journal again.

VI. CONCLUSION

IOT is a technical advancement in agriculture, where devices are in connect with farmers, which supports them technically. This this aspect, more advancement has been made and tested with different machine learning method to predict the best out of the suitable condition for better crop production. Artificial neural network has been demonstrated to be powerful tools for the prediction, to increase their effectiveness. Neural Network may validate to be an a more applicable mechanism. Considering the future scope, crop yield prediction model and deep learning algorithms have to be demonstrated for better crop prediction.

References

[1]. Hemasai Katuru, S. Ravi Kishan, Suresh Babu Dasari, " Predicting Crop Yield and Effective use of Fertilizers using Machine Learning Techniques, " International Journal of Innovative Technology and Exploring Engineering (IJITEE) May 2020, ISSN: 2278-3075, Volume-9 Issue-7.

[2]. Ashok Tatapudi, P Suresh Varma, " Prediction of Crops based on Environmental Factors using IoT & Machine Learning Algorithms, " International Journal of Innovative Technology and Exploring Engineering (IJITEE) May 2020, ISSN: 2278-3075, Volume-9 Issue-7.

[3]. Archana Gupta, Dharmil Nagda, Pratiksha Nikhare, Atharva Sandbhor, " Smart Crop Prediction using IoT and Machine Learning, " International Journal of Engineering Research & Technology (IJERT) NTASU - 2020 Conference Proceedings, ISSN: 2278-0181.

[4]. Shivani S. Kale, Preeti S. Patil, " A Machine Learning Approach to Predict Crop Yield and Success Rate, " 2019 IEEE Pune Section International Conference (PuneCon) MIT Dec 18-20, 2019.

Mr. Nilesh B. Patel completed Bachelor of commerce and Master of Computer Application from Hem. North Gujarat University, Patan (Gujarat, India). He is currently pursuing Ph.D.(in Sankalachand Patel University,Visnagar, Gujarat,India) and currently working as Assistant Professor in Shree Ambaji BCA College, Ambaji.(Department of computer science, Hem. North Gujarat University, Patan (Gujarat, India). His main research work focuses on Machine Learning. He has 8 years of teaching experience.

Dr. Mahammad Idrish I. Sandhi, completed Bachelor of commerce and Master of Computer Application and Ph. D. in Computer Science with Big Data as the major domain for research. He has 19 research papers in various conferences & Journals of national and international repute with UGC referred journals. He is recognized Ph.D. research supervisor at Sankalchand Patel University and currently guiding PhD Scholars. He has delivered many expert talks to various colleges. His few areas of research domains are Big Data, Social Media Data Analysis, e-Governance and Sentiment Analysis. He has attended more than 30 International & National, STTPs /Seminars /Workshops /Faculty Development Programs conducted