International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9, Issue-7, May 2020

Predicting Crop yield and Effective use of Fertilizers using Machine Learning Techniques

Hemasai Katuru, S. Ravi Kishan, Suresh Babu Dasari

Abstract: ***In-season crop yield estimation has various applications such as the farmer taking corrective measures to increase the yield. We are exploring the efficient use of fertilizers. The various data mining techniques are used on data for environment. The data is related to humidity, PH. value, water, soil type and atmospheric pressure these are responsible for crop yield. This result is obtained by this algorithm are useful for farmers to take decisions about further implantation of crop yield. Crop selection method is widely used to build decision tree to overcome many problems in real time real world. One of the most important fields is decision tree. By analyzing the soil and atmosphere at particular region best crop in order to have more crop yield and the net crop yield can be predicted. This prediction will help the farmers to choose appropriate crops for their farm according to the soil type, temperature, humidity, water level, spacing depth, soil PH, season, fertilizer and months. This prediction can be carried out using Random Forest classification machine learning algorithm.***

Keywords: Crop yield, Fertilizers, Humidity, Machine Learning Techniques, PH, Soil.

I. INTRODUCTION

Crop yield expectation is a significant agrarian issue. The Agricultural yield principally relies upon climate conditions, pesticides. Exact data about history of harvest yield is significant for settling on choices identified with farming danger the board and future expectations. The study of preparing machines to learn and create models for future forecasts is generally utilized, and not to no end. Agribusiness assumes a basic job in the worldwide economy. With the proceeding with extension of the human populace understanding overall harvest yield is integral to tending to nourishment security difficulties and diminishing the effects of environmental change. With the effect of environmental change in India, majority share of the agrarian yields is by and large severely influenced regarding their presentation over a time of most recent two decades. Anticipating the harvest yield well in front of its gather would help the measures for promoting and capacity. Such expectations will likewise help the related arrangement producers and ranchers for taking fitting. Enterprises for arranging the coordination's of their business. Harvest creation is a mind-boggling wonder that is impacted by climatically input parameters.

Revised Manuscript Received on May 30, 2020.
***** Correspondence Author
K. Hemasai Katuru*, CSE, V R Siddhartha Engineering College, Vijayawada, India. Email: katuruhemasai@gmail.com
S. Ravi Kishan, Computer Science and Engineering, VRSEC, Vijayawada, India. Email: suraki@vrsiddhartha.ac.in
Suresh Babu Dasari, Computer Science and Engineering, VRSEC, Vijayawada, India. Email: mdasarisuresh88@gmail.com

Agribusiness input parameters differs from field to field and rancher to rancher. Gathering such data on a bigger region is an overwhelming errand. Nonetheless, the climatic data gathered in India at each square meter territory in various pieces of the zone arranged by Indian Meteorological Department. Additionally, the yield of each harvest in each state is gathered and distributed by the branch of agribusiness and collaboration consistently. Such informational collections are utilized right now foreseeing the impact on significant harvests and along these lines, their yield in a future year.

II. RELATED WORK

Anil Suat Terliksiz et.al., concentrated on soybean yield forecast of Lauderdale County, Alabama, USA utilizing 3D CNN model that use the spatiotemporal highlights [1]. The yield is given from USDA NASS Quick Stat apparatus for a considerable length of time 2003-2016. The expectation of harvest yield has direct effect on national and worldwide economies and assume significant job in the nourishment the executives and nourishment security.

Niketa Gandhi et.al. [2] Proposed a choice emotionally supportive network model for rice crop yield forecast for Maharashtra state, India. A GUI has been made in Java utilizing NetBeans apparatus and Microsoft Office Access database for the simplicity of ranchers and leaders. The interface takes into account the determination of the scope of precipitation, least temperature, normal temperature, most extreme temperature and reference crop evapotranspiration and predicts the normal class of yield viz., low, moderate or high. Ranjini B Guruprasad et.al.,[3] introduced a contextual analysis of climate and soil information-based yield estimation demonstrating for paddy crop at various spatial goals (SR) levels, to be specific, at the area and taluk levels in India. We give a point by point investigation of precision of the yield estimation models across changed arrangements of highlights and diverse AI systems. Nilima et.al., [4] introduced a thought for example to how to send WSN on field and how Machine learning model is fitted for forecast of bug/ailments utilizing Naive Bayes Kernel Algorithm.

Remote Sensor Network is new innovation to world and nation like India where it can utilize in Agriculture Sector in India for expanding yield by giving early expectation of plant sicknesses and bug. This can be occurred by taking crude information from field where WSN organize is introduce and with fitting proper AI model for this information to get anticipated yield.Shruti Kulkarni et.al., presents a model for example an information driven model that learns by notable soil just as precipitation information to break down and anticipate crop yield over seasons in a few locales, has been created [5].

For this investigation, a specific yield, Rice is considered. The planned half breed neural system model distinguishes ideal mixes of soil parameters and mixes it with the precipitation design in a chose locale to develop the expectable harvest yield. The spine for the prescient investigation model regarding the precipitation depends on the Time-Series approach in Supervised Learning.

S. Bhanumathi et.al. Analyses the different related characteristics like area, pH esteem from which alkalinity of the dirt is resolved. Yield forecast is a significant issue in rural. Any rancher is keen on knowing how a lot of yield he is going to expect [6]. Every one of these characteristics of information will be dissected, train the information with different appropriate AI calculations for making a model.

Neha Rale et.al. [7] Propose to utilize AI procedures to build up an expectation model for crop yield creation. They analyses the exhibition of different direct and non-straight regressor models utilizing 5-overlap cross approval. Previously, ranchers used to foresee crops dependent on their own understanding and watched climate conditions. Climate, irritations, and collect activity might be kept as reference for future years.

T. Mhudchuay et.al. [8] Concentrated on downpour took care of rice where the fundamental activities are when to begin development and when to collect. The objective is to locate the ideal development and collect period to such an extent that ranchers' salary is amplified. This paper speaks to a use of a Deep Q-learning in the rice crop development practice, where the ideal activities are resolved.

Shivi Sharma et.al., [9] proposed a technique utilized, in that dirt and condition highlights for example normal temperature, normal stickiness, all out precipitation and creation yield are utilized in anticipating two classes in particular: great yield and awful yield.

Suhas S Athani et.al. [10] Presents the data relating to the harm of harvests as of late because of the development of weeds. Weeds are one of the significant hazards to the genuine home and mankind. Right now, thought, Support Vector Machine (SVM) Classifier is used to make out whether plant is harvest or weed. The maize crops are consistently observed by catching pictures utilizing camera. So as to group a plant as a yield or weed, different highlights are removed which among them are shape, surface, shading.

III. METHODOLOGY

Predicting crop yield using the powerful algorithm and determining how much fertilizer should be used to get the crop's proper yield.

In our methodology consists of following phases:

- Overview of Data
- Data Preprocessing
- Model Selection
- Crop Prediction
- Required Packages and Libraries

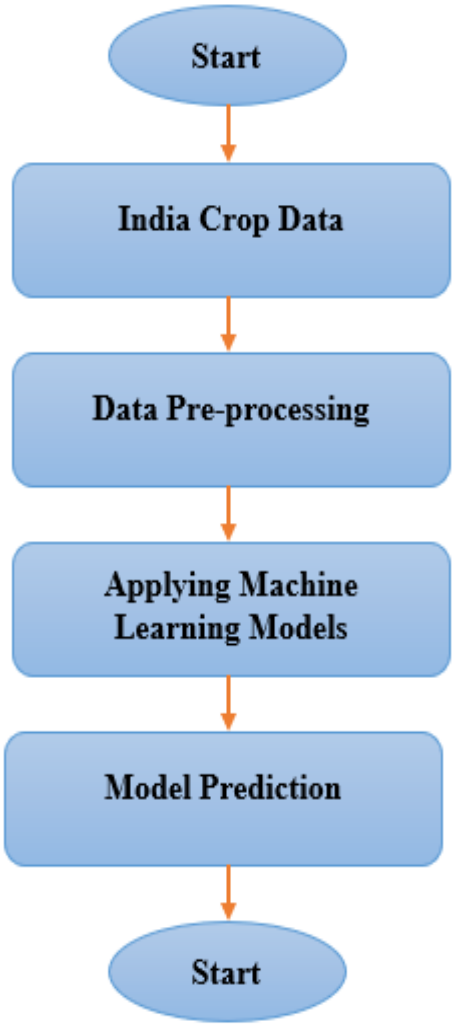

Fig 1. Process of Our Proposed Methodology

A. Overview of Data

In this paper, India crop data set was used i.e. is for prediction. This is the collection of sample data used in this project. The data used to estimate yields of crops based on 9 variables. We can construct a machine learning model and train the model using these 9 factors and we predict the output and we can predict from the data set how much fertilizer will be used to achieve the correct yield.

season	pH value	temperatu	humidity	rainfall	yield	water	crop
1	5.5	26	36	143	3.99	80	Rice
1	5.7	29	37	99	4.26	80	Rice
1	5.9	22	38	110	4.64	80	Rice
1	6	27	39	107	3.61	80	Rice
1	6.2	20	40	137	4.65	80	Rice
1	6.3	21	41	155	4.14	80	Rice
1	6.5	30	42	149	3.64	80	Rice
0	5.8	24	40	49	3.08	50	Maize
0	5.9	19	41	80	3.96	50	Maize
0	6.1	29	42	73	4.3	50	Maize
0	6.3	20	43	66	2.1	50	Maize
0	6.6	27	44	52	4.46	50	Maize
0	6.7	32	35	76	6.39	50	Maize
0	6.8	23	36	83	6.13	50	Maize
0	6	27	49	40	1.15	60	Jowar
0	6.3	24	40	50	1.33	60	Jowar
0	6.2	29	41	42	1.64	60	Jowar
0	6.5	34	42	43	1.1	60	Jowar
0	6.6	32	43	51	1.91	60	Jowar

Fig 2. Sample Image of Input Dataset 1

B. Data Pre-processing

Information Pre-preparing is a strategy that is utilized to change over the crude information into a spotless informational collection. At the end of the day, at whatever point the information is accumulated from various sources it is gathered in crude configuration which isn't doable for the investigation. Right now, information in the yield information is cleaned and the metadata is annexing to it by evacuating the things which are changed over to the whole number. Along these lines, the information is anything but difficult to prepare. Hear all the information. Right now, we first burden the metadata into this and afterward this metadata will be connected to the information and supplant the changed over information with metadata. At that point this information will be moved further and evacuate the undesirable information in the rundown and it will partition the information into the train and the test information For this parting of the information into train and test we have to import train_test_split which in the scikit-gain proficiency with this will help the pre-prepared information to part the information into train and test as indicated by the given weight given in the code. The division of the test and train is done as 20 and 80 percent individually

C. Model Selection

In this, we used different machine learning algorithms, therefore for prediction purpose we use random forest classification and logistic regression algorithm.
When preparing, each tree in a random forest gain from an irregular example of the information focuses. The examples are drawn with substitution, known as bootstrapping, which implies that a few examples will be utilized on various occasions in a solitary tree. This calculation suits for both gigantic and little information to give a productive expectation. In light of the offered information to the calculation it structures different decision trees and checks for what number of trees give a similar expectation. It depends on the votes it will check and which trees give a similar yield after that the yield given by the most extreme trees it will appear as yield as clarified in segment IV. The given information in the task go to the arbitrary woodland calculation and hear it will fabricate ten trees and pass information to it. Each tree is classified dependent on the different conditions and it will prepare the model.

Algorithm:

We actualize the working of Random Forest calculation with the assistance of following advances −
Step 1: First, start with the choice of arbitrary examples from a given dataset.
Step 2: Next, this calculation will build a decision tree for each example. At that point it will get the expectation result from each choice tree.
Step 3: In this progression, casting a ballot will be performed for each anticipated outcome.
Step 4: At last, select the most casted a ballot forecast result as the last expectation result.

D. Crop Prediction

Toward the start, the program gets to the information of the yield of that crop in that specific state over the previous years and trains that information to anticipate what the yield would be founded on the past yields. The program at that point predicts the estimations of the variables influencing the yield of the harvest in that specific year by getting to and preparing the information of elements in the state during the previous hardly any years. The results of the crop prediction obtained in the results and discussion section.

IV. RESULTS AND DISCUSSION

In this paper, exertion is made so as to know the harvest creation examination and is prepared by executing both the Random Forest algorithm and different machine learning techniques. These models were explored different avenues regarding various sorts of yields in different districts across India to foresee the yield. Indeed, even manure information was prepared utilizing these models and assessed to get the consequence of how much forecast got precisely. Both the models for the harvest creation were looked at in anticipating the yield and by different parameters concerning the mistake rate. We analyzed the mistake rate got while looking at the irregular random forest classification algorithm and remaining models where we got the blunder rate lesser to the arbitrary backwoods than different models while foreseeing the yield for both of the models. The crop data input is entered in the way and the crop data production appears in Fig. 5

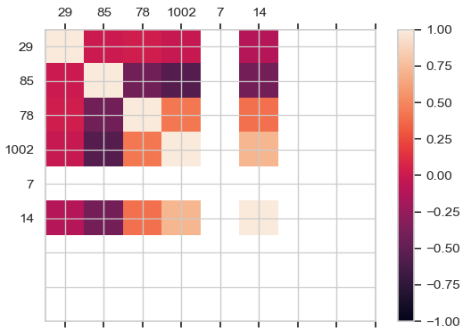

Fig 3. Correlation matrix of Data Set

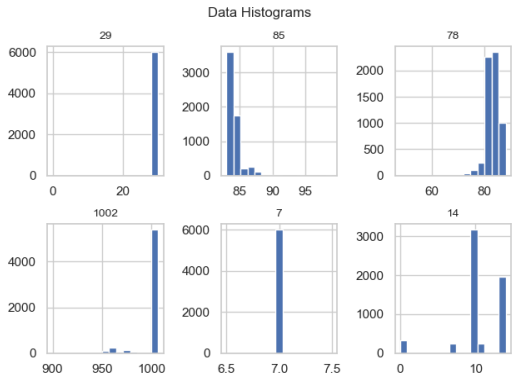

Fig 4. Data Histograms

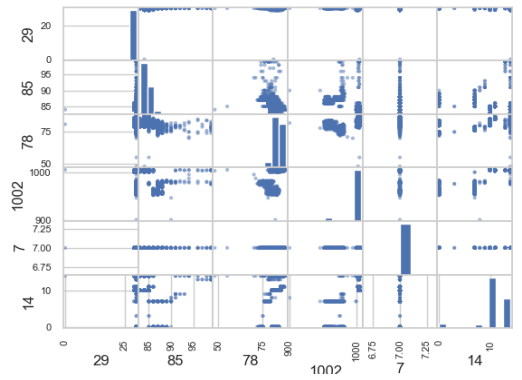

Fig 5. Scatter Matrix of Data set

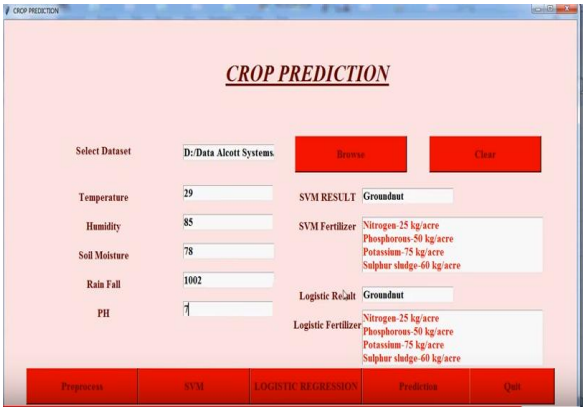

Fig 6. Crop Prediction

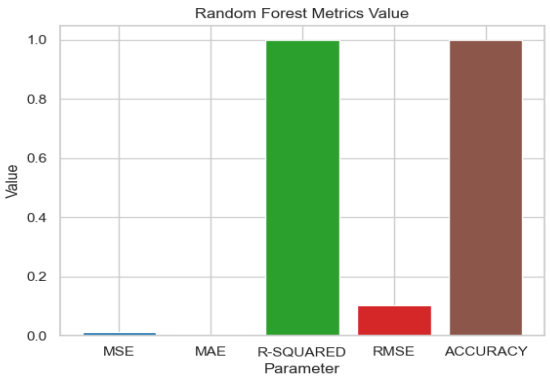

Fig 7. Performance analysis of Random Forest Algorithm

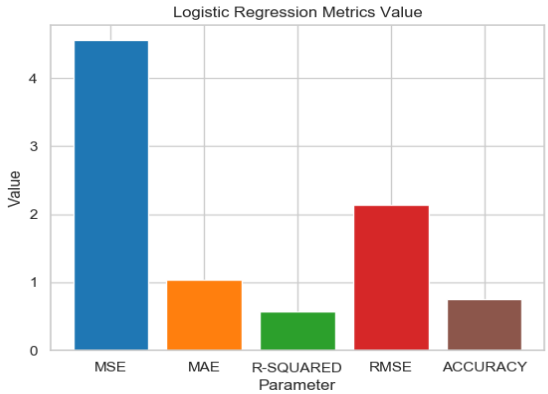

Fig 8. Performance analysis of Random Forest Algorithm

V. CONCLUSION

Crop yield forecast and proficient utilization of the compost is effectively anticipated and furthermore found the productive calculation from both the calculation and acquired the most effective yield of the yield. In future building up the web application dependent on this philosophy and make the client utilize this effectively and help the client to comprehend the yield of the harvest, he is going to trim in that season. In this, we use random forest classification algorithm and logistic regression algorithm for predicting crop prediction using India crop agriculture data set. And we get 100% model accuracy better than logistic regression. This will be particularly valuable for all farmers and related enterprises as it will assist them with planning their business early.

REFERENCES

1. Gandhi N, Armstrong LJ, Petkar O. Proposed decision support system (DSS) for Indian rice crop yield prediction. In2016 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR) 2016 Jul 15 (pp. 13-18). IEEE.
2. Terliksiz AS, Altýlar DT. Use Of Deep Neural Networks For Crop Yield Prediction: A Case Study Of Soybean Yield in Lauderdale County, Alabama, USA. In2019 8th International Conference on Agro-Geoinformatics (Agro-Geoinformatics) 2019 Jul 16 (pp. 1-4). IEEE.
3. Guruprasad RB, Saurav K, Randhawa S. Machine Learning Methodologies for Paddy Yield Estimation in India: a Case Study. InIGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium 2019 Jul 28 (pp. 7254-7257). IEEE.
4. Wani H, Ashtankar N. An appropriate model predicting pest/diseases of crops using machine learning algorithms. In2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS) 2017 Jan 6 (pp. 1-4). IEEE.
5. Kulkarni S, Mandal SN, Sharma GS, Mundada MR. Predictive Analysis to Improve Crop Yield using a Neural Network Model. In2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI) 2018 Sep 19 (pp. 74-79). IEEE.
6. Bhanumathi S, Vineeth M, Rohit N. Crop Yield Prediction and Efficient use of Fertilizers. In2019 International Conference on Communication and Signal Processing (ICCSP) 2019 Apr 4 (pp. 0769-0773). IEEE.
7. Rale N, Solanki R, Bein D, Andro-Vasko J, Bein W. Prediction of Crop Cultivation. In2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC) 2019 Jan 7 (pp. 0227-0232). IEEE.
8. Mhudchuay T, Kasetkasem T, Attavanich W, Kumazawa I, Chanwimaluang T. Rice Cultivation Planning Using A Deep Learning Neural Network. In2019 16th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON) 2019 Jul 10 (pp. 822-825). IEEE.
9. Sharma S, Rathee G, Saini H. Big data analytics for crop prediction mode using optimization technique. In2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC) 2018 Dec 20 (pp. 760-764). IEEE.
10. Athani SS, Tejeshwar CH. Support vector machine-based classification scheme of maize crop. In2017 IEEE 7th International Advance Computing Conference (IACC) 2017 Jan 5 (pp. 84-88). IEEE.

AUTHORS PROFILE

Ms. Hemasai Katuru, Contemplating MTech, Department of Computer Science and Engineering, V R Siddhartha Engineering College, Vijayawada.

S.Ravi Kishan, M.Tech (Ph.D) working as an Associate Professor in VRSEC has 10 years of research experience in the area of Data Analytics with more than 10 research publications.

Suresh Babu Dasari, M.Tech working as an Assistant Professor in VRSEC has 3 years of research experience in the area of Data Engineering.