Crop Yield Prediction Using Machine Learning

B.Manjula Josephine, K.Ruth Ramya, K.V.S.N Rama Rao, Swarna Kuchibhotla, P. Venkata Bala Kishore, S. Rahamathulla,

Abstract : Agriculture plays big role in populated countries like Bharat. AsCrop yields are basically subject to climate. A developing exact writing models this relationship so as to extend environmental change impacts on the part. We describe an approach to predict millet crop yield prediction, which can be done by taking high dimensional datasets. By using Random Forest Classifier, we obtained 99.74% of accuracy in calculating the millet crop yield prediction by taking various input fields like soil, min temp, max temp, humidity, rainfall, etc.Thus, Millet Crop Yield Prediction is an important agricultural problem this work will help the farmers to identify the crop losses and prevent it in future. We would like to extend this work to predict and find out the accuracy of millet crop yield for both Support Vector Machine(SVM) and Linear Regression (LR).

Keywords: Random Forest Algorithm; Machine Learning; Crop Yield Prediction; Millet Crop Yield; Support Vector Machine; Linear Regression.

1. INTRODUCTION

Agribusiness is that the foundation of every economy. In an exceedingly nation like Bharat, that has regularly expandinginterest of nourishment on account of rising populace, progresses in farming division territory unit expected to fulfil the necessities. From antiquated period, agribusiness is considered in light of the fact that the principle and along these lines the first culture rehearsed in Bharat. Antiquated people develop the harvests in their very own territory then they need been suited to their needs. In this way, the common yields region unit developed and is used by a few animals like characters, creatures and flying creatures. The greenish product made inside the terrains that are taken by the animal outcomes in a sound and welfare life. Since the creation of later imaginative advances and methods the horticulture field is gradually debasing. Due to these, heavy innovation people region unit been focusing on developing counterfeit item that is half and half items any place there brings about partner unfortunate life. These days, stylish people don't have mindfulness concerning the development of the yields in an exceedingly perfect time and at an opportune spot. Attributable to these developing methods the regular climatic conditions likewise are being adjusted against the essential resources like soil, water and air that reason weakness of nourishment. By examining of these issueslike climate, temperature and various different variables, there is no rightanswer and advances to beat things really young looking by United States of America. Information handling moreover supportive for foreseeing the harvest yield creation. For the most part, strategy is that the way toward breaking down information from totally various perspectives and condensing it into accommodating information. Information mining programming framework is partner scientific device that empowers clients to inquire about information from numerous different measurements or points, sort, and outline the connections known. In fact, strategy is that the way toward discovering relationships or examples among many fields in mammoth relative databases.

- *B.Manjula Josephine, K.Ruth Ramya, K.V.S.N Rama Rao, Swarna Kuchibhotla, P. Venkata Bala Kishore, S. Rahamathulla*
- *Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India.*

The examples, affiliations, or connections among this information will give information. Information will be conceived again into information concerning chronicled examples and future patterns. For instance, layout information concerning crop creation will encourage the ranchers decide the harvest misfortunes and hinder it in future. Harvest yield forecast is a significant agrarian issue. Every single rancher is generally attempts to get a handle on, what amount yield can get from his desire. Before, yield forecast was determined by breaking down rancher's past mastery on a particular harvest. The Agricultural yield is basically relies upon environmental condition, irritations and planning of reap activity.

2. RELATED WORK

Sahu, S., Chawla, M., & Khare, N. et al. [1] has dissected the harvest yield forecast system by utilizing huge information approach with the assistance of Hadoop outline work by taking the dataset from different online sources to anticipate the appropriate harvest utilizing Random Forest calculation which is incorporated with the MapReduce programming model in Hadoop structure and inferred that by utilizing this structure we can store colossal volume of yield information and can give better expectation for ranchers to plant which sort of harvests to their homesteads are reasonable dependent on the dirt substance which gives and most elevated profitability and furthermore referenced that by utilizing this Hadoop system it very well may be improved exactness in forecast. K. Samundeeswari, K. Srinivasan, et al. [2] investigates the dirt information examination for crop yield expectation in Krishnagiri District by contrasting different information mining strategies like K-Means, Support Vector Machine (SVM), and Multiple Linear Regression (MLR) which gives the greatest precision. In this paper, the significant Parameter which is utilized to expand crop creation is considered as Soil. Creators included that the work can likewise be extended and upgraded utilizing the climatic factors and harvest expectation for future purposes. Bhanumathi, S., Vineeth, M., & Rohit, N. et al. [3] has break down the different related properties like Location, esteem from which alkalinity of the dirt is resolved. Alongside it, level of supplements like Nitrogen (N), Phosphorous (P), and Potassium (K). They have taken the Location with the utilization of outsider applications like APIs for climate and temperature, kind of soil, supplement estimation of the dirt and measure of precipitation in that area, soil creation can be resolved. Every one of these properties of information

will be broke down, train the information with different appropriate AI calculations like Random Forest Algorithm (RFA), Backpropagation Algorithm for making a model. Creators Concluded that the framework accompanies a model to be exact and precise in anticipating crop yield and convey end client with appropriate proposals about required compost apportion dependent on barometrical and soil parameters of the land which improve to build the harvest yield and increment rancher income. And furthermore included that the work is stretched out for building up the web applications 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 specific season. Nevavuori, P., Narra, N., & Lipping, T et al. [4] has taken distinctive preparing calculations like Adadelta Training Algorithm, SGD-force and RMSprop and actualized utilizing Convolutional Neural Networks (CNNs) – a profound learning procedure by gathering information utilizing business off-the-rack Unmanned Aerial Vehicle (UAV) and camera bundles, and spotlight on a spatial scale that empowers to foresee intrafield yield appropriation inside the setting of individual ranch crop observing. Creators presumed that the RMSprop indicated poor combination and was in this way precluded from resulting tests. Between the two residual calculations, Adadelta outflanked SGD-force and was picked as the preparation calculation for further examinations and In future, This model will be prepared on a bigger arrangement of highlights like atmosphere and soil alongside time arrangement picture information to turn the prepared model for precision. Mohan, P., & KumariPatil, K et al. [5] has done review by taking Agricultural Statistical numeric information of Karnataka and examined that till now all trials has done either on foreseeing the reasonable climate for crop or on any of the parameters like yield or cost of the harvest. What's more, after review, Authors reasoned that if all the factual dissected yields like the climate, yield and cost of the significant harvests can be given in an independent easy to use application it will be useful for the ranchers to yield the more yield with most extreme benefits. Siju, H. L., & Patel, P. J. et al. [6] has done a Review on Crop Yield Prediction utilizing Data Mining Focusing on Groundnut Crop and Naive Bayes Technique. Right off the bat, various information mining applications in horticulture has been depicted and examined, at that point writing identified with Crop Yield Prediction has been explored alongside d Naive Bayes method actualized for various applications lastly, talked about research works for Groundnut crop yield expectation. Creators reasoned that the more exact groundnut crop yield expectation model can be developed by utilizing various information mining procedures. Mann, M. L., Warner, J. M., & Malik, A. S. et al. [7] has proposed another information combination strategy to Ethiopia which consolidates both remotely detected information (RSD) and agrarian overview information for a considerable beneficiary of specially appointed imported nourishment help. RSD is gotten close to mid-season for foreseeing significant harvest misfortunes and found at least 25% yield misfortunes at town level because of dry spell for five essential grain crops. Additionally, in the wake of establishing 81% exactness of harvest misfortunes at mid to late September creators accepted that these models can be utilized for future advancement for remotely detected information like Harmonized Landsat Sentinel (HLS) which is high goals for checking and foreseeing crop yields. Chlingaryan, A., Sukkarieh, S., & Whelan, B. et al. [8] has done an audit which is predominantly focuses and talked about on AI methods, yield estimation and, accuracy nitrogen the board. The survey exhibits the techinque of back proliferation significance and its precision of harvest yield expectation for various vegetation lists. They especially exhibits that guassian procedure are valuable for foreseeing and finding various qualities of plant leaves. Creators likewise audit the significance of M5-Prime Regression Trees are most appropriate device for finding numerous yield expectations. At last, this survey additionally exhibits on Fuzzy Cognitive Map (FCM) which will be utilized for crop yield expectation for model and portrayal of master information. Chipanshi, A., Zhang, Y., Kouadio, L., Newlands, N., Davidson, A., Hill, H., ... & Reichert, G. et al. [9] This examination depends on Integrated Canadian CropYield Forecaster (ICCYF) for three fundamental harvests at different spatial and fleeting scales shows that the exactness of the conjecture is improved over the croping season when all the more ongoing information is accessible. This gives a lead time of around multi month before har-vest and around 3–4 months before the official last arrival of the review results from Statistics Canada, which is frequently made openly accessible in December. Since the official consequences of this examination originate from the last arrival of the four yearly yield overviews. Given provincial contrasts in ICCYF quality crosswise over Canada, the discoveries were inside or superior to anything the watched yield change as detailed by Statistics Canada. It is in this way possible that, notwithstanding the may assets gave by Statistics Canada to direct four harvest yield studies during the developing season, the ICCYF may have the option to finish. Johnson, M. D., Hsieh, W. W., Cannon, A. J., Davidson, A., & Bédard, F. et al. [10] In this paper Machine learning approaches have been utilized to create crop yield forecast models for the Canadian Prairies. This investigation concentrated on determining grain, canola and spring wheat utilizing remotely detected vegetation records and thought about MODIS-NDVI, MODIS-EVI and AVHRR-NDVI viability. Mkhabela et al. (2011) results have been affirmed as it has been demonstrated that MODIS-NDVI is a successful harvest y indicator yield. In any case, it was discovered that the lower goals AVHRR-NDVI has next to zero determining abilities that are steady with the consequences of Li et al. (2007) where MODIS-NDVI was appeared to beat AVHRR-NDVI for maize and soybean yield forecast. Li et al. (2010) contrasted MODIS-NDVI and MODIS-EVI and found the NDVI out when all is said in done. The most ideal approach to bunch the CARs to create crop guaging models was not clear, as it was difficult to decide the ideal number of groups. As a rule, be that as it may, a few of the bunching calculation CAR groupings prompted models with higher abilities than those created from the CAR gathering of agro-climatic zones. On the off chance that there was a more drawn out information record, it could be feasible for us.

3. METHODOLOGY:

To find various solutions in machine learning the methodology also should be very simple.. It contains six steps in which the first step concentrates on input statistical data which is weather and forecasting crop yield dataset. The second phase, Data pre-processing techniques involves in transforming the raw data into understandable format. After pre-processing techniques, the next phase is dimensional reduction which is used to reduce the number of random variables under consideration by obtaining a set of principal variables. The entire process should undergone by using random forest algorithm in which verification of data and forecasting should be done to achieve good results.

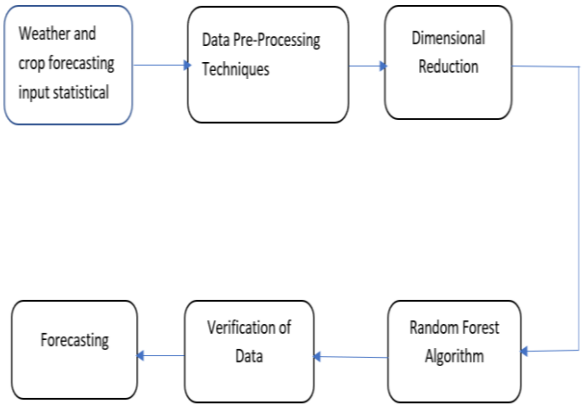


Fig. *General Forecasting block diagram*

Proposed System:

This situation mainly focuses on climate gauging, crop yield expectation and harvest cost anticipating. These elements help the ranchers to develop the best nourishment harvests and raise the correct animals byunderstanding to natural segments. Similarly, the ranchers can somewhat adapt to changes in the climate by shifting planting times, choosing assortments of different development terms or adjusting harvest pivots. The factual numerical information is identified with horticulture is embraced for the test investigation. However, the grouping based systems and the calculations administered are used to deal with the measurable information gathered. Furthermore, the reasonable grouping techniques like Random woodland (RF), Support Vector Machine (SVM), Logistic Regression (LR) and Neural Networks are utilized for better order result.

Algorithm:

Random Forest Algorithm: Random Forest Algorithm (RF) is also known as Random Forest Classifier. The following are the steps to understand RFA.

- Step-1 : Start with determining arbitrary examples from a given dataset.
- Step 2 − Next, this calculation will develop a choice tree for each example. At that point it will get the expectation result from each choice tree.
- Step 3 − In this step, voting will be performed for every predicted result.
- Step 4 –This is final step in RFA, select the most voted prediction as the final prediction result.

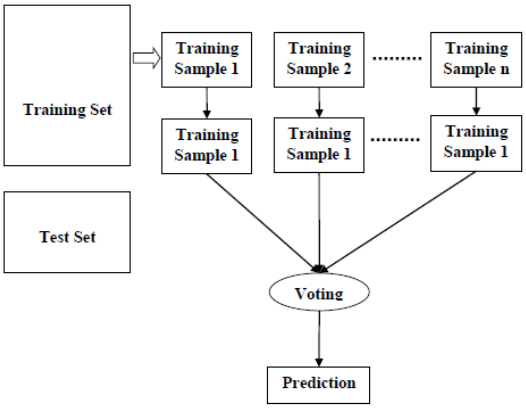


Fig. *Shows the Working scenario of Random Forest Algorithm.*

This calculation runs proficiently on huge databases and it has higher grouping precision. These systems will help in anticipating the precipitation, crop yield determining and cost forecast of harvests. Exact data about history of harvest yield is something critical for settling on choices identified with farming danger the executives. By this way, the paper proposes a plan to foresee the yield of the harvest.The rancher will check the yield of the harvest according to the section of land, before developing onto the field.

4. RESULTSAND DISCUSSION:

There are various frameworks that use different systems to control information, to determine bits of knowledge and help hesitation making for ranchers. In any case, the significant concern is that they center either around one harvest expectation or gauge anybody parameter like either yield or cost. This plan is utilized to conjecture the climate, yield and cost of significant harvests of Andhra Pradesh and Uttar Pradesh dependent on recorded information. Particularly, for Srikakulam, in light of the fact that they are the biggest maker of Millets in Andhra Pradesh. The measurable information and anticipated yield are available for the ranchers through an independent easy to understand application. This guides rancher to settle on the yield they might want to plant for the imminent year, which causes them to acquire most extreme cost for their items.

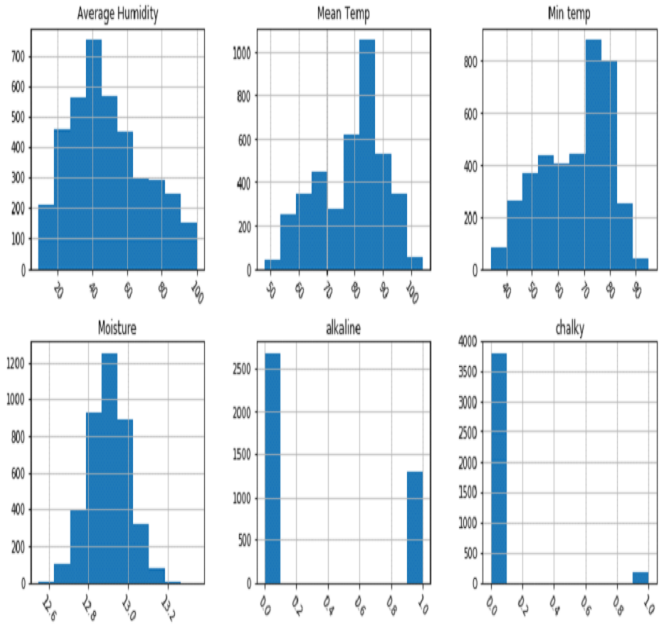
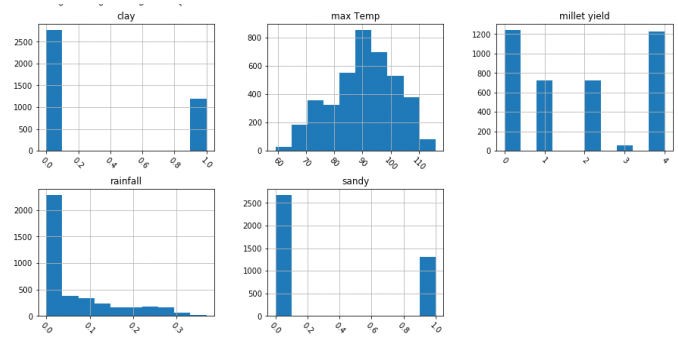


Fig. shows the histogram graphs of various fields like Average Humidity, Mean Temp, Min Temp, Moisture, Alkaline, Chalky. All these fields are taken as raw input data sets. Each histogram graph has 10 bins. The input raw datasets also consists of various fields like Dry, Max Temp, Millet Yield, Rainfall and, Sandy as shown below.



CONCLUSION:

We can infer from the above graphs that when the alkaline is minimum (0.0) the millet yield is maximum (2500). And when the chalky content is low (0.0) the millet yield is maximum (3500). And when chalky content is maximum then millet yield is minimum. Next, moisture content is average then weobtain maximum yield for millet. Also, In this paper we have applied Random Forest Algorithm and predicted the Millet crop yield in which it has getting 99.74% of accuracy by taking statistical raw datasets which contains various fields likeAverage Humidity, Mean Temp, Min Temp, Moisture, Alkaline, Chalky, Dry, Max Temp, Millet Yield, Rainfall and, Sandy. As python is easy which has run time results and Jupyter is useful for showing different graphs the entire implementation is done by using Anaconda Software which consists of both Python IDLE and Jupyter Notebook. By importing Random Forest Classifier to our datasets which will be test and train we are successfully able to predict the millet crop yield prediction for the taken datasets. The outcomes are as shown in below,

Random Forest Classifier

```
In [61]: from sklearn.tree import DecisionTreeClassifier
         from sklearn.ensemble import RandomForestClassifier

In [62]: rf = RandomForestClassifier(n_estimators=140)
         rf.fit(X_train,n_train)
         print (rf.score(X_test,n_test))

         0.997487437186
```

REFERENCES:

[1] Sahu, S., Chawla, M., & Khare, N. (2017, May). An efficient analysis of crop yield prediction using Hadoop framework based on random forest approach. In 2017 International Conference on Computing, Communication and Automation (ICCCA) (pp. 53-57). IEEE.

[2] K. Samundeeswari, K. Srinivasan, “CROP YIELD PREDICTION AND SOIL DATA ANALYSIS USING DATA MINING TECHNIQUES IN KRISHNAGIRI DISTRICT”, International Journal of Computer Sciences and Engineering, Vol.06, Special Issue.08, pp.49-55, 2018.

[3] Bhanumathi, S., Vineeth, M., & Rohit, N. (2019, April). Crop Yield Prediction and Efficient use of Fertilizers. In 2019 International Conference on Communication and Signal Processing (ICCSP) (pp. 0769-0773). IEEE.

[4] Nevavuori, P., Narra, N., & Lipping, T. (2019). Crop yield prediction with deep convolutional neural networks. Computers and Electronics in Agriculture, 163, 104859.

[5] Mohan, p., & kumariPatil, K. (2017). Survey on Crop and Weather Forecasting based on Agriculture related Statistical Data. International Journal of Innovative Research in Computer and Communication Engineering, 5(2).

[6] Siju, H. L., & Patel, P. J. (2018). Review on Crop Yield Prediction using Data Mining Focusing on Groundnut Crop and Naive Bayes Technique.

[7] Mann, M. L., Warner, J. M., & Malik, A. S. (2019). Predicting high-magnitude, low-frequency crop losses using machine learning: an application to cereal crops in Ethiopia. Climatic Change, 154(1-2), 211-227.

[8] Chlingaryan, A., Sukkarieh, S., & Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. Computers and electronics in agriculture, 151, 61-69.

[9] Chipanshi, A., Zhang, Y., Kouadio, L., Newlands, N., Davidson, A., Hill, H., ... & Reichert, G. (2015). Evaluation of the Integrated Canadian Crop Yield Forecaster (ICCYF) model for in-season prediction of crop yield across the Canadian agricultural

landscape. Agricultural and Forest Meteorology, 206, 137-150.

[10] Johnson, M. D., Hsieh, W. W., Cannon, A. J., Davidson, A., & Bédard, F. (2016). Crop yield forecasting on the Canadian Prairies by remotely sensed vegetation indices and machine learning methods. Agricultural and forest meteorology, 218, 74-84.

[11] Rao, K. R., & Josephine, B. M. (2018, October). Exploring the Impact of Optimal Clusters on Cluster Purity. In 2018 3rd International Conference on Communication and Electronics Systems (ICCES) (pp. 754-757). IEEE.