INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE

ISSN: 2057-5688

Crop Yield Prediction Using Machine Learning

Chittimella Swetha
B-Tech Student

Ayyagari Srinivasa Sai Phani Chnadra
B-Tech Student

Bhardwaj Pramod
B-Tech Student

Arfa Mahvish
(Assistant Professor)

Department of Information Technology
CMR Technical Campus
Kadlakoya (V), Medchal, Hyderabad-501401

Abstract: Agriculture growth mainly depends on diverse soil parameters, namely Nitrogen, Phosphorus, Potassium, Crop rotation, Soil moisture, pH, surface temperature and weather aspects like temperature, rainfall, etc. Technology will prove to be beneficial to agriculture which will increase crop productivity resulting in better yields to the farmer. The proposed project provides a solution for Smart Agriculture by monitoring the agricultural field which can assist the farmers in increasing productivity to a great extent. This work presents a system, in a form of a website, which uses Machine Learning techniques in order to predict the most profitable crop in the current weather and soil conditions. This system can also help in predicting the yield of the crop using weather parameter, soil parameter and historic crop yield. Thus, the project develops a system by integrating data from various sources, data analytics, prediction analysis which can improve crop yield productivity and increase the profit margins of farmer helping them over a longer run.

I. INTRODUCTION

Agriculture is the backbone of India's economy since its plays a vital role in the survival of every human and animal in India. The worldwide population was estimated at 1.8 billion in 2009 and is predicted to increase to 4.9 billion by 2030, leading to an extreme increase in demand for agricultural products. In the future, agricultural products will have higher demand among the human population, which will require efficient development of farmlands and growth in the yield of crops. Meanwhile, due to global warming, the crops were frequently spoiled by harmful climatic situations . A single crop failure due to lack of soil fertility, climatic variation, floods, lack of soil fertility, lack of groundwater and other such factors destroy the crops which in turn affects the farmers. In other nations, the society advises farmers to increase the production of specific crops according to the locality of the area and environmental factors. The population has been increasing at a significantly higher rate, so the estimation and monitoring of crop production is necessary. Accordingly, an appropriate

method needs to be designed by considering the affecting features for the better selection of crops with respect to seasonal variation. The core objective of crop yield estimation is to achieve higher agricultural crop production and many established models are exploited to increase the yield of crop production. Nowadays, ML is being used worldwide due to its efficiency in various sectors such as forecasting, fault detection, pattern recognition, etc. The ML algorithms also help to improve the crop yield production rate when there is a loss in unfavourable conditions. The ML algorithms are applied for the crop selection method to reduce the losses crop yield production irrespective of distracting environment. The existing model used SVM that classified the crop data based on the texture, shape, color of patterns on the diseased surface as it includes an unambiguous perception of the defects. An existing technique used CNN that reduced

the relative error as well as decreased the prediction of crop yield . Similarly, the existing model used Back Propagation Neural Network (BPNNs) with the time series model and used smaller dataset size gained lower performance as less number of sample was used for prediction. Crop yield prediction using machine learning There are various machine techniques used in agriculture for yield prediction, smart irrigation system, Crop disease prediction, crop selection, weather forecasting, deciding the minimum support price, etc. These techniques will enhance the productivity of the fields along with a reduction in the input efforts of the farmers. Besides, the advances in machines and technologies were accurate as they used significant data and played an important role. This research work analyses the various agricultural methods that utilize ML, along with the merits and limitations.

II. IMPLEMENTATION

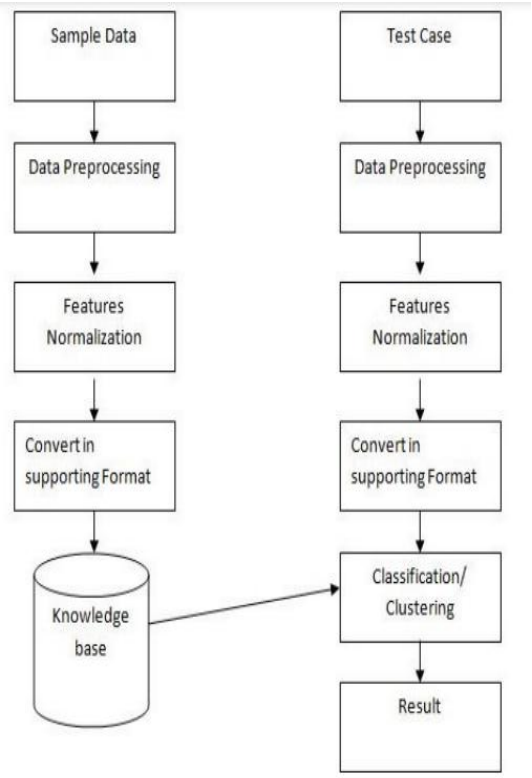

FIG-1: System Architecture

Methodology

Data Collection

Data is composed from a different source and optimized for data sets. And the data is used to evaluate descriptively. Several abstract online outlets, like Kaggle, Google weather forestation and data government, provide the data for up to 10years in series. The data sets such as soil nature, climatic conditions and seed data are used for the crop prediction and better crop yields.

Data Pre-processing

Pre-processing the data is considered as a significant step machine learning phase. Pre-processing involves adding the missing values, the correct set of data, and extracting the functionality. Data set form is important to the process of analysis. The data collected in this step will induced in Google Colab platform in the form of python programming in order to get the desired output.

Feature Extraction

Extraction of the features would reduce the data size involved to characterize a wide collection of data. The characteristics of soil, crop and weather collected from the pre-treatment process establish the final training data collection.

Crop Prediction

In advance to this step there need to split the data into train dataset and test dataset. By applying the Naïve Bayes Gaussian classifier the data is trained with available input and output data. In the test phase, the data are tested if the accuracy of e model is satisfied. Then the new data is predicted by machine learning module.

Yield Prediction

This is a kind of ensembling but a little of enhancement of averaging. In this, we add a Meta model and use the out of fold predictions of the other models used to train the main Meta model.

- Step-1: the total training set is again divided into two different sets. (Train and holdout)
- Step-2: train the selected base models with first part (train).
- Step-3: Test them with the second part. (holdout)

- Step-4: Now, the predictions obtained from test part are inputs to the train higher level learner called meta-model.

Iteratively, the first three steps are completed. For example, if we take a 5-fold stacking, we divide the training data into 5 folds first. We'll then do 5 iterations. We train each base model on 4 folds in each iteration and predict the remaining fold (holdout fold). So, after 5 iterations, we'll be confident that all the data will be used to get out - of-fold predictions that we'll use as a new feature in Step 4 to train our meta-model. We average the predictions of all base models on the test data for the predictive portion and used them as meta-features on which the meta-model is finally predicted.

Yield Forecasting

Forecasting is the process of predicting the future using current and previous data. The major challenge to understanding the patterns in the sequence of data and then using this pattern to analyze the future. If we were to hand-code the patterns, it would be tedious and changes for the next data. Deep Learning has proven to be better in understanding the patterns in both structured and unstructured data. To understand the patterns in a long sequence of data, we need networks to analyze patterns across time. Recurrent Networks is the one usually used for learning such data. They are capable of understanding long and short term dependencies or temporal differences.

Random Forest

The random forest is an ensemble approach that can also be thought of as a form of nearest neighbour predictor. Ensembles are a divide-and-conquer approach used to improve performance. The main principle behind ensemble methods is that a group of 'weak learners' can come together to form a 'strong learner'. The random forest starts with a standard machine learning technique called a 'decision tree' which, in ensemble terms, corresponds to our weak learner. The decision tree algorithm repeatedly splits the data set according to a criterion that maximizes the separation of the data, resulting in a tree-like structure. In this algorithm an input is entered at the top and as it traverses down the tree the data gets bucketed into smaller and smaller sets. The random forest takes this notion to the next level by combining trees with the notion of an ensemble. Thus, in ensemble terms, the trees are weak learners and the random forest is a strong learner. The advantages of a random forest classifier are that its' runtimes are quite fast, and that it is able to deal with unbalanced and missing data. Weaknesses of this algorithm are that when used for regression it cannot predict beyond the range in the training data, and it may over-fit data sets that are particularly noisy.

III. RESULT

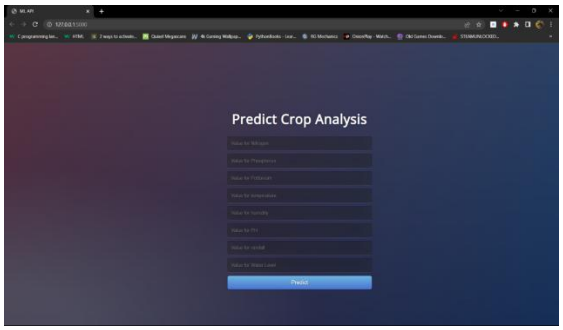

Fig-7.3: Values Entry Screen

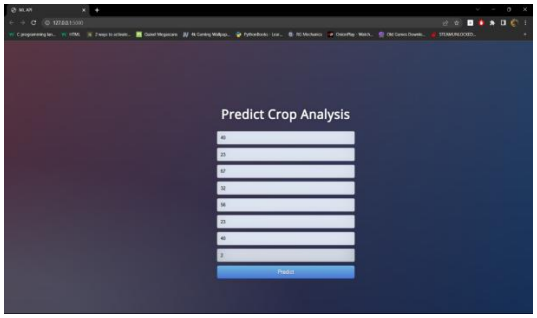

FIG-7.4 Crop Prediction Screen

IV. CONCLUSION

The Results shows that we can attain an accurate crop yield prediction using the Random Forest algorithm. Random Forest algorithm achieves a largest number of crop yield models with lowest models. It is suitable for massive crop yield prediction in agricultural planning. This makes the

farmers to take the right decision for right crop such that the agricultural sector will be developed by innovative ideas. Agriculture is the field which helps in economic growth of our country. But this is lacking behind in using new technologies of machine learning. Hence our farmers should know all the new technologies of machine learning and other new techniques. These techniques help in getting maximum yield of crops. Many techniques of machine learning are applied on agriculture to improve yield rate of crops. These techniques also help in solving problems of agriculture. We can also get the accuracy of yield by checking for different methods. Hence we can improve the performance by checking the accuracy between different crops. Sensor technologies are implemented in many farming sectors. This paper helps in getting maximum yield rate of the crops. Also helps in selecting proper crop for their selected land and selected season. These techniques will solve the problems of farmers in agriculture field. This will help in improving the economic growth of our country.

REFERENCES

1. Hyndman, R. et al., 2008. Forecasting with Exponential Smoothing: The State Space Approach, Springer Science & Business Media.
2. Box, G.E.P. et al., 2015. Time Series Analysis: Forecasting and Control, John Wiley & Sons.
3. Box, G.E.P. & Cox, D.R., 1964. An Analysis of Transformations. Journal of the Royal Statistical Society. Series B, Statistical methodology, 26(2), pp.211252.
4. Yeo, J. et al., 2016. Browsing2purchase: Online Customer Model for Sales Forecasting in an E-Commerce Site. 25th Int.Conf.Comp. on World Wide Web.
5. Ramanathan, U., 2013. Supply chain collaboration for improved forecast accuracy of promotional sales. Int. Journal of Operations & Production Management.
6. Kulkarni, G., Kannan, P.K. & Moe, W., 2012. Using online search data to forecast new product sales. Decision support systems, 52(3), pp.604611.
7. Zhao, K. & Wang, C., 2017. Sales Forecast in E-commerce using Convolutional Neural Network. arXiv [cs.LG].
8. Seeger, M.W., Salinas, D. &Flunkert, V., 2016. Bayesian Intermittent Demand Forecasting for Large Inventories. In Proceedings of the 29th NIPS.
9. Snyder, R., Ord, J.K. & Beaumont, A., 2012. Forecasting the intermittent demand for slow-moving inventories: A modelling approach. IJF, 28(2), pp.485496.
10. Zhang, G., Patuwo, B.E. & Hu, M.Y., 1998. Forecasting with artificial neural networks: The state of the art. International journal of forecasting, 14(1), pp.3562.
11. Yan, W., 2012. Toward automatic time-series forecasting using neural networks. IEEE transactions on neural networks and learning systems, 23(7), pp.10281039.
12. Zimmermann, H.-G., Tietz, C. & Grothmann, R., 2012. Forecasting with RNNs: 12 Tricks. In Neural Networks. Lecture Notes in Computer Science. pp. 687707.
13. Trapero, J.R., Kourentzes, N. & Fildes, R., 2015. On the identification of sales forecasting models in the presence of promotions. The Journal of the ORS.
14. Borovykh, A., Bohte, S. &Oosterlee, C.W., 2017. Conditional Time Series Forecasting with Convolutional Neural Networks arXiv [cs.AI].
15. Flunkert, V., Salinas, D. &Gasthaus, J., 2017. DeepAR: Probabilistic Forecasting with Autoregressive Recurrent Networks. arXiv [cs.AI].