CROP YEILD PREDICTION BASED ON INDIAN AGRICULTURE USING MACHINE LEARNING

Mr. N.Kesava Rao , M.Tech(Ph.D), Assoc.Professor Department Of Computer Science and Engineering, Narayana Engineering College (Autonomous), Gudur,AP, India
D. Nandhini, CH. Hyndhavi, K.Sowmya, UG Students, Department of Computer Science and Engineering, Narayana Engineering College (Autonomous College-Gudur, AP, India

Abstract *– This project predicts the yield of almost all kinds of crops that are planted in India. This script makes novel by the usage of simple parameters like State, district, season, area and the user can predict the yield of the crop in which year he or she wants to. The paper uses advanced classification techniques like Random forest, Decision Tree and Gradient Boost algorithms to predict the yield and uses the concept of Stacking Regression for enhancing the algorithms to give a better prediction.*
Index terms – Crop yield prediction, Random Forest, Decision tree, Gradient Boost

INTRODUCTION

In our research, which we found in the previous research papers is that everyone uses climatic factors like rainfall, sunlight and agricultural factors like soil type, nutrients possessed by the soil (Nitrogen, Potassium, etc.) but the problem is we need to gather the data and then a third party does this prediction and then it is explained to the farmer and this takes a lot of effort for the farmer and he doesn't understand the science behind these factors. To make it simple and which can be directly used by the farmer this paper uses simple factors like which state and district is the farmer from, which crop and in what season (as in Kharif, Rabi, etc.). In India, there are more than a hundred crops planted around the whole country. These crops are categorized for better understanding and visualization.

The data for this research has been acquired from the Indian Government Repository [1]. The data consists of attributes – State, District, Crop, Season, Year, Area and Production with around 2.5 Lakh observations. The fig.1.depicts the states and territories of India which visualize that which category of crops are famous in which season. We used advanced regression techniques – Lasso, ENet and Kernel Ridge and further we used stacking of these models to minimize the error and to obtain better predictions[2][3].

LITERATURE SURVEY

An improved crop yield prediction model using bee hive clustering approach for agricultural data sets Agricultural researchers over the world insist on the need for an efficient mechanism to predict and improve the crop growth[4]. The need for an integrated crop growth control with accurate predictive yield management methodology is highly felt among farming community. The complexity of predicting the crop yield is highly due to multi dimensional variable metrics and unavailability of predictive modeling approach, which leads to loss in crop yield. This research paper suggests a crop yield prediction model (CRY) which works on an adaptive cluster approach over dynamically updated historical crop data set to predict the crop yield and improve the decision making in precision agriculture. CRY uses bee hive modeling approach to analyze and classify the crop based on crop growth pattern, yield[5]. CRY classified dataset had been tested using Clementine over existing crop domain knowledge.

An intelligent system based on kernel methods for crop yield prediction

This paper presents work on developing a software system for predicting crop yield from climate and plantation data. At the core of this system is a method for unsupervised partitioning of data for finding spatio-temporal patterns in climate data using kernel methods which offer strength to deal with complex data[7][8]. For this purpose, a robust weighted kernel k-means algorithm incorporating spatial constraints is presented. The algorithm can effectively handle noise, outliers and auto-

correlation in the spatial data, for effective and efficient data analysis, and thus can be used for predicting oil-palm yield by analyzing various factors affecting the yield.

Fuzzy Logic based Crop Yield Prediction using Temperature and Rainfall parameters predicted through ARMA, SARIMA, and ARMAX models.

Agriculture plays a significant role in the economy of India. This makes crop yield prediction an important task to help boost India's growth. Crops are sensitive to various weather phenomena such as temperature and rainfall. Therefore, it becomes crucial to include these features when predicting the yield of a crop. Weather forecasting is a complicated process. In this work, three methods are used to forecast- ARMA (Auto Regressive Moving Average), SARIMA (Seasonal Auto Regressive Integrated Moving Average) and ARMAX (ARMA with exogenous variables). The performance of the three is compared and the best model is used to predict rainfall and temperature which are in turn used to predict the crop yield based on a fuzzy logic model.

Crop Yield Prediction Using Data Analytics and Hybrid Approach

Agricultural data is being produced constantly and enormously. As a result, agricultural data has come in the era of big data. Smart technologies contribute in data collection using electronic devices. In our project we are going to analyze and mine this agricultural data to get useful results using technologies like data analytics and machine learning and this result will be given to farmers for better crop yield in terms of efficiency and productivity[6].

A study on various data mining techniques for crop yield prediction.

India is a country where agriculture and agriculture related industries are the major source of living for the people. Agriculture is a major source of economy of the country. It is also one of the country which suffer from major natural calamities like drought or flood which damages the crop. This leads to huge financial loss for the farmers thus leading to the suicide.

Predicting the crop yield well in advance prior to its harvest can help the farmers and Government organizations to make appropriate planning like storing, selling, fixing minimum support price, importing/exporting etc. Predicting a crop well in advance requires a systematic study of huge data coming from various variables like soil quality, pH, EC, N, P, K etc. As Prediction of crop deals with large set of database thus making this prediction system a perfect candidate for application of data mining[9][10].

PROPOSED SYSTEM

In this section we the system model of the project in figure 1.

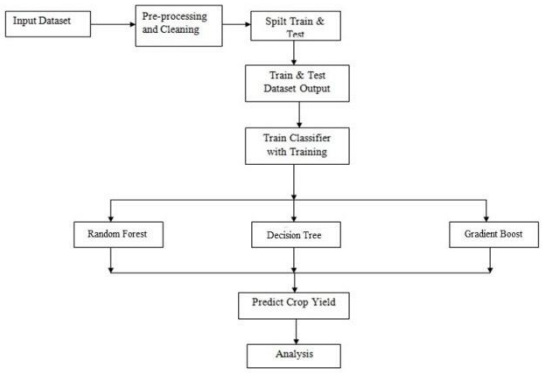

Fig. 1: System Overview

Implementation Modules

Preprocessing

For the given data set, there are quite a few ‘NA’ values which are filtered in python. Furthermore, as the data set consists of numeric data, we used robust scaling, which is quite similar to normalization, but it instead uses the interquartile range whereas normalization is something which normalization shrinks the data in terms of 0 to 1.

Stacked Regression

- 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.

Graphical Analysis

In this phase of the Implementation user can get the clear picture analysis of the cause of death analysis. Various factors take into consideration for the graph analysis. In this phase plot the charts like pie graph, bar.

Implementation Algorithms

This project used classification technique for prediction. Algorithm has been selected by evaluating each supervised machine learning technique. The target of classification is to predict future event by each classifier. In this work four classifiers are used namely Random Forest, Decision Tree Regression and Gradient Boosting Regression. The prediction result of all classifiers are analyzed and compared.

RANDOM FOREST REGRESSOR

- It generates multi decision trees from which each decision tree uses a part of data sample and predicts the result.
- Then the result which was achieved by maximum number of trees is considered as the final prediction.
- Random forest is a Supervised Learning algorithm which uses ensemble learning method for classification and regression. Random forest is a bagging technique and the trees in random forests run in parallel without any interactions.
- A Random Forest operates by constructing several decision trees during training time and outputting the mean of the classes as the prediction of all the trees.

DECISION TREE REGRESSION

- Trees are constructed through an algorithmic approach that identifies ways to split the data set based on different conditions.
- It is one of the most widely used practical methods for supervised learning.
- These are non-parametric method used for both classification and regression.

GRADIENT BOOST REGRESSION

- Gradient boosting method converts the weak learners into strong learners by boosting their capability.
- A sequential process of learning from the previous trees and increasing the model accuracy.
- One of the most used and efficient method.

CONCLUSION

When we apply stacked regression, the result has been so improvised than when those models were applied individually. The output which has been shown in figure is currently a web application, but our future work would be building an application where the farmers can use it as app and converting the whole system in their regional language.

REFERENCES

[1] “data.gov.in.” [Online]. Available: https://data.gov.in/

[2] Ananthara, M. G., Arunkumar, T., & Hemavathy, R. (2013, February). CRY—an improved crop yield prediction model using bee hive clustering approach for agricultural data sets. In 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering (pp. 473-478). IEEE.

[3] Awan, A. M., & Sap, M. N. M. (2006, April). An intelligent system based on kernel methods for crop yield prediction. In Pacific-Asia Conference on Knowledge Discovery and Data Mining (pp. 841-846). Springer, Berlin, Heidelberg.

[4] Bang, S., Bishnoi, R., Chauhan, A. S., Dixit, A. K., & Chawla, I. (2019, August). Fuzzy Logic based Crop Yield Prediction using Temperature and Rainfall parameters predicted through ARMA, SARIMA, and ARMAX models. In 2019 Twelfth International Conference on Contemporary Computing (IC3) (pp. 1-6). IEEE.

[5] Bhosale, S. V., Thombare, R. A., Dhemey, P. G., & Chaudhary, A. N. (2018, August). Crop Yield Prediction Using Data Analytics and Hybrid Approach. In 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA) (pp. 1-5). IEEE.

[6] P.Venkateswara Rao, Dr.A.RamamohanReddy, Dr.V.Sucharita Advances in Machine Learning Techniques for Penaeid Shrimp Disease Detection: A Survey,.International Journal of Engineering and Applied Sciences (IJEAS) ISSN: 2394-3661, Volume-3, Issue-8, August 2016.

[7] Gandhi, N., Petkar, O., & Armstrong, L. J. (2016, July). Rice crop yield prediction using artificial neural networks. In 2016 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR) (pp. 105-110). IEEE.

[8] Gandhi, N., Armstrong, L. J., Petkar, O., & Tripathy, A. K. (2016, July). Rice crop yield prediction in India using support vector machines. In 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE) (pp. 1-5). IEEE.

[9] Gandhi, N., Armstrong, L. J., & Petkar, O. (2016, July). Proposed decision support system (DSS) for Indian rice crop yield prediction. In 2016 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR) (pp. 13-18). IEEE.

[10] Islam, T., Chisty, T. A., & Chakraborty, A. (2018, December). A Deep Neural Network Approach for Crop Selection and Yield Prediction in Bangladesh. In 2018 IEEE Region 10 Humanitarian Technology Conference (R10-HTC) (pp. 1-6). IEEE.

[11] M. Venkatesan, C. S. Subash Kumar, M. Sathiyanathan (2020) A Step by Step Design of PV System: A Research Approach for Engineering Graduates (Vol. 16. No-1,), 30-35