A Deep Neural Network Approach for Classification and Recommendation of Crops in Karnataka

L.Divya
Dept. of CSE
Ramaiah Institute of Technology
Bengaluru, India
sln.divya123@gmail.com

Summaiya Sarvari V K
Dept. of CSE
Ramaiah Institute of Technology
Bengaluru, India
summaiya.sarvari.v.k@gmail.com

Aishwarya Sitharam
Dept. of CSE
Ramaiah Institute of Technology
Bengaluru, India
aishu97.s@gmail.com

Sini Anna Alex
Dept. of CSE
Ramaiah Institute of Technology
Bengaluru, India
sinialex@msrit.edu

***Abstract*— In Indian Economy, Agriculture plays an important role. Many rural households, above 70% of them are dependent on agriculture for their livelihood. Agriculture makes vital contributions to Indian economy, about 17% to the total GDP. Having a prior idea of crop yield can help the farmer in making decision about the right crop to be cultivated. This paper presents a system that classifies a crop as high yield crop and low yield crop. The system considers various factors such as district, season, crop, etc. for classification of crops. The system uses Artificial Neural Network (ANN) which is a framework for Machine Learning algorithms. ANN mimics human brain in learning and is capable of adapting to changes in itself in response to the changes in the environment. The system also aims at providing a forecast of the crop yield using Facebook's Prophet package. Prophet is different from traditional ARIMA model in that it does not require expertise in forecasting. ARIMA requires lag order, degree of differencing and order of moving average to be given as input. But adjusting these parameters for right behavior requires expertise. Prophet makes intuition on parameters which makes forecasting easier. The system also gives recommendations on the crop that can be cultivated in a particular place and a given season. The recommendation is made using Random forest algorithm and Adaptive Boosting.**

Keywords*— *Artificial Neural Network (ANN), Prophet, Machine Learning, Random forest, Adaptive Boosting.

I. INTRODUCTION

Agriculture, the primary occupation of a developing economy like India, serves as the backbone of Indian economy. The need to maintain the sustainability in agriculture is high. Technology, with its enormous growth, has paved a way for the monitoring and managing different factors of agriculture. Precision agriculture (PA) is one among the technologies. Precision agriculture (PA) is a method to farm management using technology pertaining to information(IT). It is to satisfy the demand of optimum health and productivity of all crops and a healthy yield. The goal of Precision Agriculture is to ensure profitability, sustainability and protection of the environment pertaining to agriculture.

The prediction of yield i.e.., the production of crop yield is one of the main intensions of this paper. To achieve this, the main prediction is done by employing deep learning algorithm such as artificial neural network. This algorithm helps predicting the yield by considering various factors of the agricultural system. Also, the prophet methodology is also used for forecasting the time series data on additive model.

Deciding on the crop to be grown in the region, which gives good produce and helps farmers in getting profits for their produce, can be a difficult decision for the farmer. Prediction of the crops that can be grown in that particular region, in that season, will help those who practice agriculture to take a decision on which crop he will cultivate. Planting the right crops in the right place helps the farmers get profits out of their produce. This tool is also intended on predicting the crops that can be grown in that concerned region and thus help out the farmers with the major part of decision making of cultivation. Fig 1 is the overall diagram depicting the three modules of the system through which classification of yield, forecasting of

yield for periods specified and recommendation of crops based on the input.

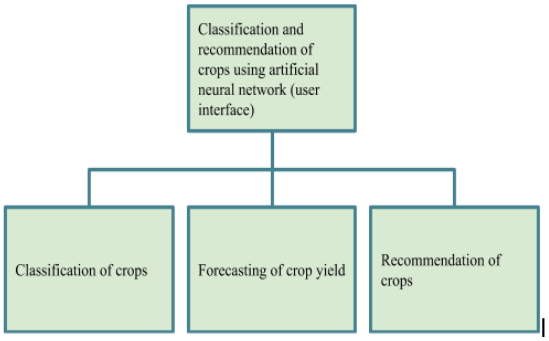

Fig. 1. Overall diagram depicting system modules

II. RELATED WORK

The research by W. Fan, C. Chong, G. Xiaoling, Y. Hua and W. Juyun [1] is about predicting crop yield using big data analytics. The paper overcomes the failure of previous algorithm to handle massive data by using big data. The proposed system makes use of nearest neighbor along with Map Reduce for classification.

The research by Z. Hong, Z. Kalbarczyk and R. K. Iyer [2] is about data driven methodology for developing precision agriculture(PA) solution. Machine learning model like Support Vector Machine (SVM) and Relevance Vector Machine (RVM) is used which makes prediction on soil moisture n days ahead.

The research by M. Paul, S. K. Vishwakarma and A. Verma [3] is about prediction of crop yield by transforming into classification rule where naïve Bayes and K nearest neighbors methods are used.

The research by S. Dimitriadis and C. Goumopoulos [4] uses machine learning techniques which automatically extracts new knowledge in the form of decision rules. The study results in the generation of effective set of decision rules that are used to predict plant's state and prevent unpleasant impact of water stress in plants. The paper proposes the use of neural network and genetic algorithms in black box approach.

The research by G. Yi-yang and R. Nan-ping [5] proposes the use of decision tree for classification. Clustering technique method is adopted to discretize the continuous data. Decision trees are the best suited because it transfers information into rules, it is crucial for decision.

The research by H. Li, Z. Chen, W. Wu, Z. Jiang, B. Liu and T. Hasi [6] proposes the use of data assimilation method which combines crop growth model with multi source observed This method is the most effective method to simulate crop growth process and predict crop yield.

The research by Arooj A et al. [7] have proposed the recommendation of crops based on the soil properties in different agricultural regions by applying various data mining techniques like Naïve Bayes and BF tree. The suggestion is that recommendation of crops leads to not only increase in yield but also reduced reliance on fertilizers.

The research by Ponce-Guevara et al. [8] use a software to provide a predictive model for soil moisture. Big Data and Data Mining techniques applied on vegetable crops data help in analyzing the factors which influence crop growth in a greenhouse. These are the main factors that affect crop growth. For the purpose C4.5 is used and visually analyzed.

The research by Kulkarni et al. [9] propose a model which learns from historic soil and rainfall data to predict the yield of crops over seasons in several districts. The designed hybrid neural network model identifies patterns in that region to predict the crop yield. Time series model is applied for prediction with respect to rainfall.

The research by Sujatha et al. [10] has imbibed the classification using Support Vector Machine (SVM), J48 algorithm, Decision tree and Random Forest for prediction of crop yield. This work also contributes the selection of best crop by the farmer based on the weather conditions.

The research by Y. Gandge and Sandhya [11] analyses the accuracy of various data mining techniques like multiple linear regression, support vector machine, neural network , C4.5, etc on predicting crop yield. The study shows the accuracy of neural network on corn yield as 95%.

The research by T. Islam, T. A. Chisty and A. Chakrabarty [12] uses backpropagation to train deep neural network with 3 hidden layers to select crops and to predict crop yield. This paper makes a prediction on crop yield of six major crops in Bangladesh Aus rice, Aman rice, Boro rice, Jute, Wheat, and Potato.

The research by Thirumalai et al. [13] proposes a model on rainfall which is one of the factors for agriculture. The rate of rainfall in each season is

considered in order to predict the rainfall. This is thus helpful for farmers and easier to practice agriculture, where the prediction is performed using linear regression.

The research by Zhao et al. [14] proposes how big data and data mining techniques can be used in the production of wheat crop. Information about the wheat crop growth, knowledge about the conditions appropriate for its growth, biochemical agents used, pests and rodents control knowledge are stored and used for making intelligent agricultural decisions. The model gives intelligent decisions in control and prevention of disease, decision of variety, sowing time decision, planting quantity and many such important decisions helpful for crop growth.

The research by Tlhobogang et al. [15] proposes a model to deal with plant diseases and their management on diagnosis. This is to help in better agricultural practices and help farmers to take appropriate measures when their plants have diseases. The model involves use of image analysis and conventional neural networks and also machine learning algorithms such as supervised learning for classification. On these results, the application gives tips, chemical and biological, to control and prevent such diseases.

III. METHODOLOGY

A. Artificial Neural Network

The deep learning technique of Artificial Neural Networks (ANN) is used for classification of the crop's yield. The factors for classification are the District_Name, Crop_Year, Season, Crop, Area and the classification of crop is defined by the Production field. In this methodology, we employ the ANN technique, considering 3 layers, namely, the input layer, middle layer and output layer. The back-propagation technique of neural networks is used to train and model the epochs. By label-encoding the predictors of classification, the algorithm successfully builds a model through which prediction is done.

The back-propagation algorithm of ANN is described as follows:

The first step involves the computation of the Δ values for the output units, using the observed error. Then, we start with the output layer and repeat the same for all the layers in the network, till the earliest hidden layer is reached:

a) Propagate the Δ values back to the previous layer.

b) Update the weights between the two layers.

B. Facebook's Prophet

Forecasting of crop yield is very essential and helpful tool for farmer since it provides the farmer with prior knowledge of the crop yield. The proposed methodology makes use of Facebook's Prophet which is an open source software. Prophet forecasts time series data based on additive model. In the proposed methodology, the dataset on which forecast is done is a time series yearly data with the crop yield in tonnes per hectare. The dataset contains the yearly data for the required crop and season and district which is based on the input. The output consists of the forecasted crop yield for the specified number of periods along with uncertainty levels.

C. Recommendation of Crops

Recommendation of crops is a helpful tool which provides a solution to an important part of the process of cultivation. Random Forest is an ensemble method which can be used for both classification as well as regression. It consists of several decision trees which output their results and the majority of these is provided as the result of the application of the algorithm. It is the most probable result and also prevents overfitting. Thus, considered as a suitable algorithm for many predictions.

The accuracy of the prediction obtained is improved by using a boosting ensemble model, Adaptive Boosting (AdaBoost). AdaBoost is a model which learns from previously made mistakes in its classification. It increases the weights of previously misclassified points and gives the weighted majority as the final prediction.

The user provides inputs about the district and the season in which he prefers to grow crops. The Random Forest algorithm is applied on the available data, which ranges over several years and with the necessary data. Then, the majority answer from the various decision trees built on the data, is chosen as the final result of the algorithm. This result is the recommended crop to the user, which is to be grown in that particular region and in that season. The accuracy obtained is then improved by the application of the AdaBoost Classification algorithm. Here, the improvement is with the addition of weights, which the algorithm follows and implements. The weighted majority from the trees constructed is provided as the final prediction. This result is the recommended crop to the user, which is to be grown in that particular region and in that season.

IV. ALGORITHMIC DETAILS

A. Classification of Crops using ANN

Following is pseudocode:

```
Begin
//Import the dataset
data=read("Crop_Production.csv")
//Adding class labels
if production<threshold:
    class=0 // low yield crop
else:
    class=1 // high yield crop
set x= predictors
set y=response
encode the categorical variables
split the dataset into training and test dataset
//creation of artificial neural network
create_ann(hidden_layer=1,neurons=3,optimizer=Ada
 m, activation= relu,sigmoid)
//training
    train the ANN on the training data
//testing
    test the ANN on the test data
calculate the accuracy
//Input
input (district name)
input (season)
input (crop name)
class=predict(new crop)
return class
End
```

B. Forecasting Crop Yield using Facebook 's Prophet

Following is the pseudocode:

```
Begin
//Import the time series dataset
data=read("Crop_Production.csv")
//Input
input (district name)
input (season)
input (crop name)
//Input number of periods
input(n)
df = subset dataset according to input
forecast = prophet(df,n)
return forecast
```

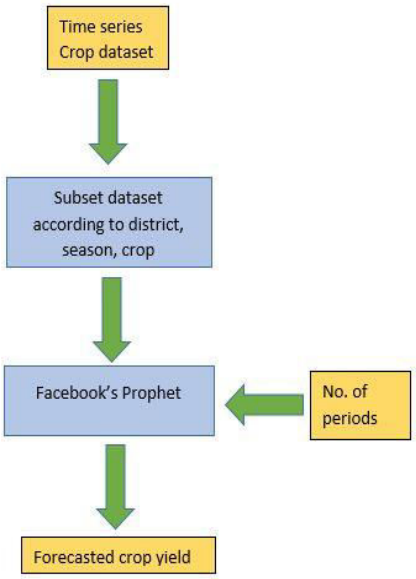

Fig. 2. Overview of forecasting of crop yield

Fig 2 represents the overall architectural view of the process of forecasting the crop yield. The time series crop dataset is a yearly dataset on the crop yield in various states and districts. Only the subset of dataset with state Karnataka is considered for forecasting. Further the subset of dataset as per the input from the farmer is considered. Facebook's prophet forecasting algorithm is applied on the dataset. The algorithm requires the number of periods for which the forecasting has to be done as input. The output consists of forecasted crop yield for the mentioned number of periods.

C. Recommendation of Crop using Random Forest and Adaptive Boosting

Following is the pseudocode:

```
Begin
//Import the time series dataset
data=read("kar_crops.csv")
//Input
input (district name)
input (season)
set x= predictors
set y=response
//split into train and test data
 rf=RandomForestClassifier(with suitable number of
    estimators)
//training
Apply Random Forest on the training data
//testing
```

Apply Random Forest on the test data
calculate the accuracy
//Input
input (district name)
input (season)
crop=predict()
return crop

ab=AdaBoostClassifier(with suitable number of estimators)
//training
Apply AdaBoost on the training data
//testing
Apply AdaBoost on the test data
//improved accuracy observed
calculate the accuracy
//Input
input (district name)
input (season)
crop=predict ()
return crop
End

V. RESULTS

	Actual	Predicted
0	0	[False]
1	0	[False]
2	0	[False]
3	0	[False]
4	0	[False]
5	0	[False]
6	1	[True]
7	0	[False]
8	1	[True]
9	1	[True]
10	0	[False]
11	0	[False]
12	1	[True]
13	0	[False]
14	1	[True]

Fig. 3. Classification results by ANN

```
print(metrics.accuracy_score(y_test,y_pred))
```

```
0.8804733727810651
```

Fig. 5. Accuracy of classification of crops

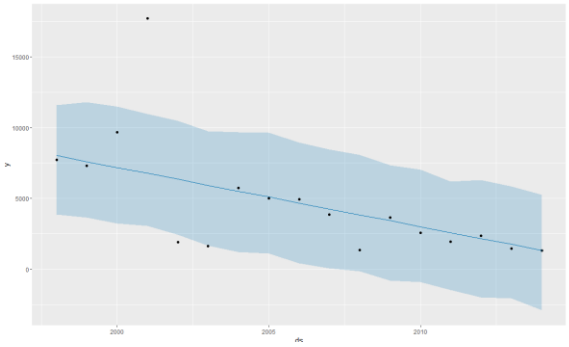

Fig. 6. Visualization of forecast

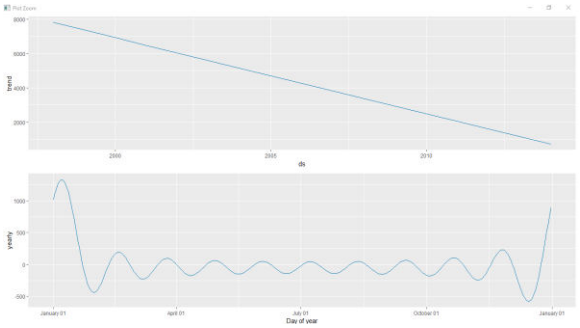

Fig. 7. Trend and yearly seasonality

Fig 3 shows the output of classification of crops by using artificial neural network.
Fig 4 shows the accuracy of classification of crops by ANN. This shows that the proposed system gives good results.
Fig 6 is the visualization of forecast of crop yield.
Fig 7 shows the trend in the time series dataset as captured by Prophet. The plot shows a linear downward trend. This shows that the system have captured the correct trend in the crop yield as the crop yield has a downward trend.
Fig 8 shows the actual and forecasted values. The forecasting model have performed well which can be seen from the difference between the actual and forecasted values.

Filter

	Date	Actual	Predicted	y_lower	y_upper
1	1998-01-01	7734.000	7660.0999	3712.1626	11797.662
2	1999-01-01	7326.000	7224.8773	3055.3634	11382.987
3	2000-01-01	7439.701	6789.7010	2432.7354	10765.002
4	2001-01-01	17691.000	6403.9681	2369.2536	10217.342
5	2002-01-01	6618.715	5968.7146	1801.5231	10253.355
6	2003-01-01	6183.489	5533.4888	1607.2883	9526.264
7	2004-01-01	5748.309	5098.3093	1196.4228	9568.328
8	2005-01-01	5000.000	4712.5732	658.6494	8729.415
9	2006-01-01	4927.317	4277.3166	269.5672	7997.978
10	2007-01-01	3866.000	3842.0907	-121.2035	7607.630
11	2008-01-01	4056.911	3406.9112	-821.3512	7700.003
12	2009-01-01	3671.175	3021.1751	-1020.1385	7334.644
13	2010-01-01	2564.000	2585.9184	-1582.5533	6536.407
14	2011-01-01	1945.000	2150.6926	-1750.0942	6366.828
15	2012-01-01	2365.513	1715.5131	-2516.4487	5700.167
16	2013-01-01	1454.000	1329.7770	-3039.5217	5189.737
17	2014-01-01	1344.000	894.5203	-2890.7292	5169.743

Fig. 8. Actual versus Forecasted crop yields.

Fig. 9. Recommendation of Crop by Random Forest and AdaBoost.

Fig 9 shows that, the recommended crop is the crop suitable for cultivation in that particular district and season. The recommendation has been done using Random Forest and AdaBoost classification algorithms.

VI. CONCLUSION AND FUTURE WORK

The tool helps out farmers by predicting the yield which can be gained by cultivation of that particular crop in that region. It provides accurate results of the same, as the prediction is done with the data ranging over many years. Also providing with predictions of yields of various such crops as required by the people practicing agriculture. The tool also suggests the crop on prediction performed on the basis of the region and the crop seasons. Thus, helping in the process of decision making of the crop that can be grown in that region to get good produce.

A suggestion of fertilizers for the crops grown would be more helpful to the farmers. The usage of real-time data in addition to the already available existing data would help in real-time analysis and prediction, thus, helping the farmers get the latest updates.

REFERENCES

[1] W. Fan, C. Chong, G. Xiaoling, Y. Hua and W. Juyun, "Prediction of Crop Yield Using Big Data," *2015 8th International Symposium on Computational Intelligence and Design (ISCID)*, Hangzhou, 2015, pp.

[2] Z. Hong, Z. Kalbarczyk and R. K. Iyer, "A Data-Driven Approach to Soil Moisture Collection and Prediction," *2016 IEEE International Conference on Smart Computing (SMARTCOMP)*, St. Louis, MO, 2016, pp.

[3] M. Paul, S. K. Vishwakarma and A. Verma, "Analysis of Soil Behaviour and Prediction of Crop Yield Using Data Mining Approach," *2015 International Conference on Computational Intelligence and Communication Networks (CICN)*, Jabalpur, 2015, pp.

[4] S. Dimitriadis and C. Goumopoulos, "Applying Machine Learning to Extract New Knowledge in Precision Agriculture Applications," *2008 Panhellenic Conference on Informatics*, Samos, 2008, pp.

[5] G. Yi-yang and R. Nan-ping, "Data mining and analysis of our agriculture based on the decision tree," *2009 ISECS International Colloquium on Computing, Communication, Control, and Management*, Sanya, 2009, pp.

[6] H. Li, Z. Chen, W. Wu, Z. Jiang, B. Liu and T. Hasi, "Crop model data assimilation with particle filter for yield prediction using leaf area index of different temporal scales," *2015 Fourth International Conference on Agro-Geoinformatics (Agro-geoinformatics)*, Istanbul, 2015, pp.

[7] A. Arooj, M. Riaz and M. N. Akram, "Evaluation of predictive data mining algorithms in soil data classification for optimized crop recommendation," *2018 International Conference on Advancements in Computational Sciences (ICACS)*, Lahore, 2018, pp.

[8] K. L. Ponce-Guevara *et al*., "GreenFarm-DM: A tool for analyzing vegetable crops data from a greenhouse using data mining techniques (First trial)," *2017 IEEE Second Ecuador Technical Chapters Meeting (ETCM)*, Salinas, 2017, pp.

[9] S. Kulkarni, S. N. Mandal, G. S. Sharma, M. R. Mundada and Meeradevi, "Predictive Analysis to Improve Crop Yield using a Neural Network Model," *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Bangalore, 2018, pp.

[10] R. Sujatha and P. Isakki, "A study on crop yield forecasting using classification techniques," *2016 International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE'16)*, Kovilpatti, 2016, pp. [11] Y. Gandge and Sandhya, "A

study on various data mining techniques for crop yield prediction," *2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT)*, Mysuru, 2017, pp.

[12] T. Islam, T. A. Chisty and A. Chakrabarty, "A Deep Neural Network Approach for Crop Selection and Yield Prediction in Bangladesh," *2018 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*, Malambe, Sri Lanka, 2018, pp.

[13] C. Thirumalai, K. S. Harsha, M. L. Deepak and K. C. Krishna, "Heuristic prediction of rainfall using machine learning techniques," *2017 International Conference on Trends in Electronics and Informatics (ICEI)*, Tirunelveli, 2017, pp.

[14] B. Tlhobogang and M. Wannous, "Design of plant disease detection system: A transfer learning approach work in progress," *2018 IEEE International Conference on Applied System Invention (ICASI)*, Chiba, 2018, pp.

[15] J. Zhao and J. Guo, "Big data analysis technology application in agricultural intelligence decision system," *2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*, Chengdu, 2018, pp.