IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

AGRICULTURE PRODUCTION PREDICTION USING MACHINE LEARNING ALGORITHMS

[1]D Mohan Reddy, [2]Dr.T.Manikumar, [3]Dr.R.Maruthamuthu

[1]PG Research Scholar, [2, 3]Assistant Professor.

[1, 2, 3] Department of Computer Applications,

[1, 2, 3] Madanapalle Institute of Technology and Science, Madanapalle, Andhra Pradesh, India, 517352.

Abstract: In general, agriculture is the backbone of India and also plays an important role in Indian economy. Most of Indians have agriculture as their occupation. Farmers usually have the mind-set of planting the same crop, using more fertilizers and following the public choice. But now-a-days, food production and prediction is getting depleted due to unnatural climatic changes, which will adversely affect the economy of farmers by getting a poor yield and also help the farmers to remain less familiar in getting high yield of crops. Machine learning is one such advanced technique deployed to predict crop yield in agriculture. By looking at the past few years, there have been significant developments in how machine learning can be used in various industries and research.

The surveyed research papers have given a rough idea about using ML with only one attribute. It gives us an idea for the finest predicted crop which will be cultivate in the field weather conditions [1]. These predictions can be done by a machine learning algorithm called Random Forest. Various machine learning techniques we are using such as KNN, DT, RF, BAGGING and GRADIENT BOOSTING [7]. However, the selection of the appropriate algorithm from the pool of available algorithms imposes challenge to the researchers with respect to the chosen crop. This system will be useful to justify which crop can be grown in a particular region [8].

Key words - **Machine Learning, Data Prediction, DT, RF, KNN, BAGGING and GRADIENT BOOSTING.**

I. INTRODUCTION

Crop production may be a complicated development that's influenced by soil and environmental condition input parameters [4]. Agriculture input parameters vary from field to field and farmer to farmer [2]. Collection such info on a bigger space may be a discouraging task. There are completely different foretelling methodologies developed and evaluated by the researchers everywhere the globe within the field of agriculture or associated sciences [4]. A number of such studies are: Agricultural researchers in alternative countries have shown that tries of crop yield maximization through pro-pesticide state policies have LED to hazardously high chemical usage [15]. These studies have reported a correlation between chemical usage and crop yield [1]. Agriculture is associate trade sector that's benefiting powerfully from the event of detector technology, knowledge science, and machine learning (ML) techniques within the latest years [3] [6]. Most of the work tired the sector of yield foretelling via cubic centimetre makes use of some kind of remote sensing knowledge over the farm. Agriculture seeks to extend and improve the crop yield and therefore the quality of the crops to sustain human life.

Later on, by victimization computer code supported machine learning, one will timely assess the temperature change impact and check attainable situations that incorporate ascertained changes in climatic conditions and water distribution. data {processing} is that the process of analysing the experimental knowledge collected over a amount and varied locations from completely different views, extract trends or patterns {of data of knowledge of info} and switch them into helpful information for users [12]. Users will then additionally reason and/or summarize the relationships ascertained from the collected knowledge, and typically predict what knowledge to expect [10]. This information provided by machine learning will facilitate farmers with crop cultivation by predicting probabilities of crop [7].

AN OVERVIEW OF MACHINE LEARNING

Machine learning: Terminology and Definition ML methodologies typically involve a learning process with the goal of learning to perform a task from "experience" (training data). In machine learning, data is made up of examples [1]. Individual examples are typically described by a set of attributes, also known as features or variables [15]. Nominal (enumeration), ordinal (e.g., A+ or B-), binary (i.e., 0 or 1), or numeric features are all possible (integer, real number, etc.)[5][13]. A performance metric that improves with experience is used to assess the ML model's performance in a specific task. Various statistical and mathematical models are used to calculate the performance of ML models and algorithms. The trained model can then be used to classify, predict, or cluster new examples (testing data) based on the experience gained during the training process [7]. Figure 1 depicts a typical machine learning approach.

II LITERATURE REVIEW

Literature survey is the most important step in software development process. Before developing the tool it is necessary to determine the time factor, economy and company strength. Once these things are satisfied, ten next steps are to determine which operating system and language can be used for developing the tool. Once the programmers start building the tool the programmers need lot of external support. This support can be obtained from senior programmers, from book or from websites. Before building the system the above consideration are taken into account for developing the proposed system. [1][2][3][5]

With the impact of climate change in India, majority of the agricultural crops are being badly affected interms of their performance over a period of last two decades. Predicting the crop yield well ahead of its harvest would help the policy makers and farmers for taking appropriate measures for marketing and storage. Such predictions will also help the associated industries for planning the logistics of their business. Several methods of predicting and modeling crop yields have been developed in the past with varying rate of success, as these don't take into account characteristics of the weather, a n d are mostly empirical. In the present study a software tool named `Crop Advisor' has been developed as an user friendly web page for predicting the influence of climatic parameters on the crop yields.C4.5 algorithm is used to find out the most influencing climatic parameter on the crop yields of selected crops in selected districts of Madhya Pradesh. This software provides an indication of relative influence of different climatic parameters on the crop yield, other agro-input parameters responsible for crop yield are not considered in this tool, since, application of these input parameters varies with individual fields in space and time.[20][21][23]

As we are aware of the fact that, most of Indians have agriculture as their occupation. Farmers usually have the mindset of planting the same crop, using more fertilizers and following the public choice. By looking at the past few years, there have been significant developments in how machine learning can be used in various industries and research. So we have planned to create a system where machine learning can be used in agriculture for the betterment of farmers. The surveyed research papers have given a rough idea about using ML with only one attribute. We have the aim of adding more attributes to our system and ameliorate the results, which can improve the yields and we can recognize several patterns for predictions. This system will be useful to justify which crop can be grown in a particular region. [15][16][17]

Yield forecast is essential to agriculture stakeholders and can be obtained with the use of machine learning models and data coming from multiple sources. Most solutions for yield forecast rely on NDVI (Normalized Difference Vegetation Index) data, which is time-consuming to be acquired and processed. To bring scalability for yield forecast, in the present paper we describe a system that incorporates satellite-derived precipitation and soil properties datasets, seasonal climate forecasting data from physical models and other sources to produce a pre-season prediction of soybean/maize yield---with no need of NDVI data. This system provides significantly useful results by the exempting the need for high-resolution remote-sensing data and allowing farmers to prepare for adverse climate influence on the crop cycle. In our studies, we forecast the soybean and maize yields for Brazil and USA, which corresponded to 44% of the world's grain production in 2016. Results show the error metrics for soybean and maize yield forecasts are comparable to similar systems that only provide yield forecast information in the first weeks to months of the crop cycle.[18][19][20][21].

Agriculture has been the sector of paramount importance as it feeds the country population along with contributing to the GDP. Crop yield varies with a combination of factors including soil properties, climate, elevation and irrigation technique. Technological developments have fallen short in estimating the yield based on this joint dependence of the said factors. Hence, in this project a data-driven model that learns by historic soil as well as rainfall data to analyse and predict crop yield over seasons in several districts, has been developed. For this study, a particular crop, Rice is considered. The designed hybrid neural network model identifies optimal combinations of soil parameters and blends it with the rainfall pattern in a selected region to evolve the expectable crop yield. The backbone for the predictive analysis model with respect to the rainfall is based on the Time-Series approach in Supervised Learning. The technology used for the final prediction of the crop yield is again a branch of Machine Learning, known as Recurrent Neural Networks. With two inter-communicating data-driven models working at the backend, the final predictions obtained were successful in depicting the interdependence between soil parameters for yield and weather attributes.

Agriculture is the main occupation of India. More than 70% of the population is involved in agriculture and its ancillary. In order to feed the expanding population there is a need to incorporate the latest technologies and tools in the agriculture sector. With the help of big data analytics, IoT, and machine learning algorithms the crop productivity can be increased by many folds. Big data provide facilities like data storage, data processing, and data analysis with accuracy, hence its use in the field of agriculture can benefit farmers and nation's economic growth. In this work, a precision agriculture model is presented to suggest farmers, which crop to cultivate according to field conditions. Focusing mainly on the agriculture in Telangana region, the model uses a Naïve Bayes classifier to recommend about the crop to the farmers. It also suggests which crop can be grown in a specific given environment.

III PROPOSED METHODOLOGY:

In this project we are performing crops prediction for district level [3]. So main aim is to find the dataset which contains production details and also details about climatic parameters and soil parameters like rainfall, temperature, moisture, etc. details. These factors will help in the prediction of the crops by using various regressor algorithms on the given dataset [6]. Thus, various factors are assessed and the factors strongly leading to accurate prediction of the crops yields production. We tested decision tree, naïve Bayes, and Random forest, bagging and gradient boosting with sample dataset [11].

Advantages:

1. Improved performance.
2. Less cost effective.
3. High efficiency.
4. Accurate accuracy.

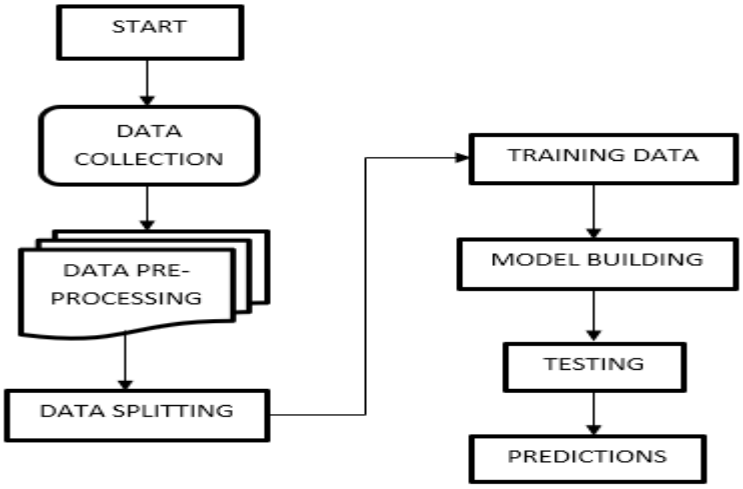


Figure3.1: Block Diagram

ARCHITECTURE

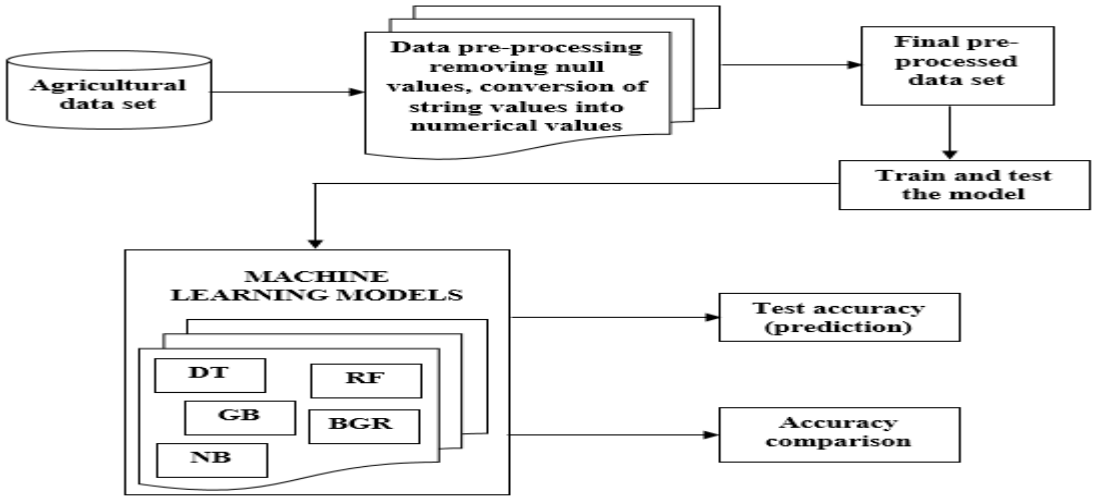


Figure 3.2: Architecture

IV RESULTS & DISCUSSION

- Firstly, we have collected Crops related dataset [7].
- Later we will load the collected dataset to our working environment [6].
- Necessary pre-processing steps will be completed here before building our required model [12].
- Dividing the data into train and test splits [18].
- Perform building machine learning model in a flask environment using Python [15].
- The model has been built with Lasso and Ridge regressions [20].
- Designed as, the system delivers the prediction results to the user depending on the inputs entered [9].

MODULES

Upload Dataset:

First we prepare the dataset and then we can upload the dataset and the data set to view the data set in library.

View:

Upload the dataset that data can be view.

Preprocessing:

Preprocessing the data the data will be extract and remove the empty files and to clean the null values and dataset. Converting string values into numerical values for further process.

Regression:

We can classify the data the data will be classified into the target value and performance the result values and based to use different types of algorithm wise to performance the Machine learning and to target value and predict the values. Algorithm they are used should show the baste results.

Graph:

Graph will be viewed by the crops to predict for the graph.

4.1 DECISION TREE:

Decision trees are non-parametric supervised learning Method used for classification and regression [6]. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features [8] [10].

A decision tree is drawn upside down with its root at the top. In the image on the left, the bold text in black represents a condition/internal node, based on which the tree splits into branches/ edges [24]. The end of the branch that doesn't split anymore is the decision/leaf, in this case, whether the passenger died or survived, represented as red and green text respectively.

Block Diagram for Decision Tree Algorithm:

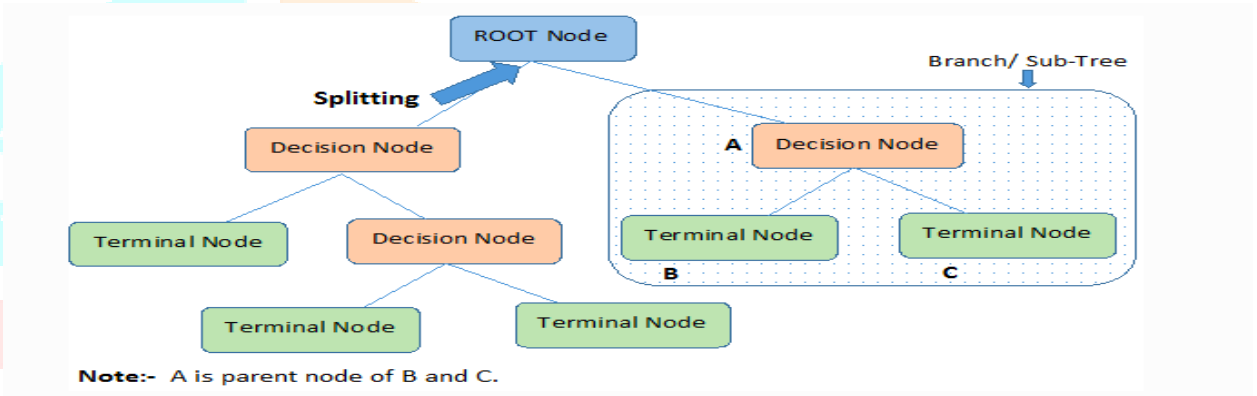


Figure 4.1: Block Diagram for Decision Tree Algorithm.

Root Node: It represents the entire population or sample and this further gets divided into two or more homogeneous sets.

Splitting: It is a process of dividing a node into two or more sub-nodes.

Decision Node: When a sub-node splits into further sub-nodes, then it is called the decision node.

Leaf / Terminal Node: Nodes do not split is called Leaf or Terminal node.

Pruning: When we remove sub-nodes of a decision node, this process is called pruning. You can say the opposite process of splitting.

Branch / Sub-Tree: A subsection of the entire tree is called branch or sub-tree.

4.2 RANDOM FOREST

A large number of relatively uncorrelated models (trees) operating as a committee will outperform any of the individual constituent models. The low correlation between models is the key [12] [20]. Just like how investments with low correlations (like stocks and bonds) come together to form a portfolio that is greater than the sum of its parts, uncorrelated models can produce ensemble predictions that are more accurate than any of the individual predictions [21].

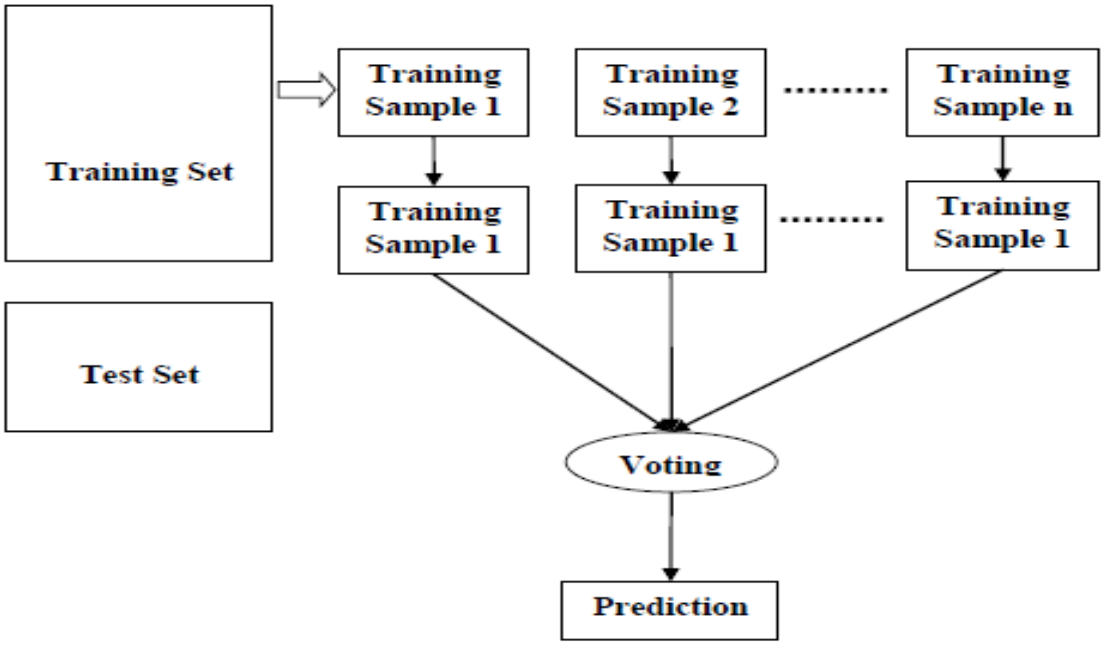


Figure 4.2: Block Diagram for Random Forest Algorithm.

4.3 NAIVE BAYES ALGORITHM:

It is a classification technique based on Bayes' Theorem with an assumption of independence among predictors. In simple terms, a **Naive Bayes classifier** assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature [7] [24].

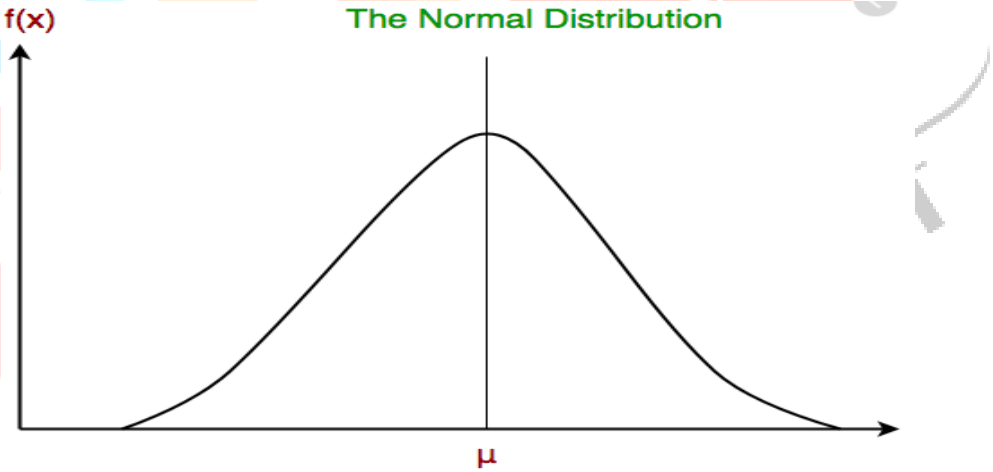
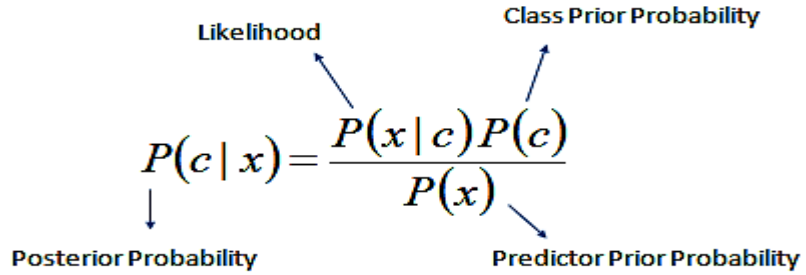


Figure 4.3: Block Diagram for Naïve Bayes Algorithm.

Naive Bayes methods are a set of supervised learning algorithms based on applying Bayes' theorem with the "naive" assumption of conditional independence between every pair of features given the value of the class variable [22].



$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

4.4 GAUSSIAN NAÏVE BAYES:

Gaussian Naive Bayes implements the Gaussian Naive Bayes algorithm for classification. The likelihood of the features is assumed to be Gaussian [19] [22].

$$P(x_i \mid y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}\right)$$

The parameters σ_y and μ_y are estimated using maximum likelihood.

4.5 GRADIENT BOOSTING

Boosting is an ensemble modeling technique which attempts to build a strong classifier from the number of weak classifiers [7]. It is done building a model by using weak models in series. Firstly, a model is built from the training data. Then the second model is built which tries to correct the errors present in the first model [9] [11]. This procedure is continued and models are added until either the complete training data set is predicted correctly or the maximum number of models are added.

AdaBoost was the first really successful boosting algorithm developed for the purpose of binary classification. ***AdaBoost*** is short for ***Adaptive Boosting*** and is a very popular boosting technique which combines multiple "weak classifiers" into a single "strong classifier" [14].

Figure 4.4: AdaBoost Algorithm.

This is another very popular Boosting algorithm which works pretty similar to what we've seen for AdaBoost. Gradient Boosting works by sequentially adding the previous predictors underfitted predictions to the ensemble, ensuring the errors made previously are corrected [16].

The difference lies in what it does with the underfitted values of its predecessor [17]. Contrary to AdaBoost, which tweaks the instance weights at every interaction, this method **tries to fit the new predictor to the *residual errors* made by the previous predictor.**

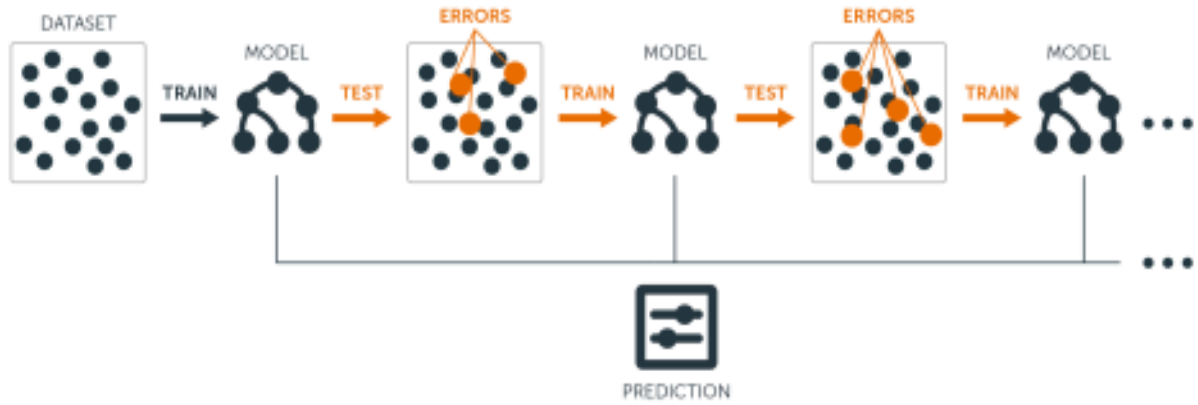


Figure 4.5: Block Diagram for AdaBoost

4.6 BAGGING REGRESSOR

A **Bagging regressor** is an ensemble meta-estimator that fits base **regressors** each on random subsets of the original dataset and then aggregate their individual predictions (either by voting or by averaging) to form a final prediction[23][19].

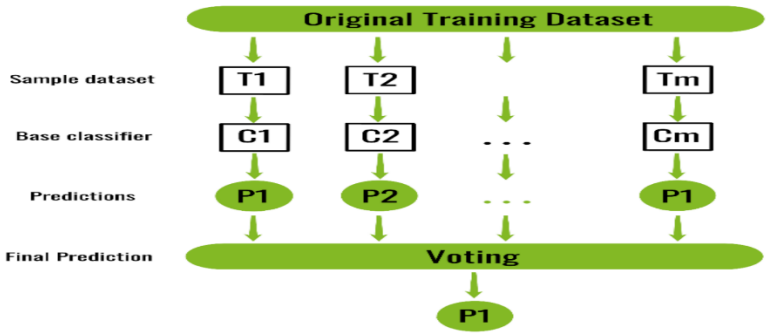


Figure 4.6: Block Diagram for Bagging Regressor

View Upload Dataset

	District_Name	Moisture	rainfall	Crop	Average Humidity	Mean Temp	Season	amt of n	amt of p	amt of k	Production	Area	Status_of_y
0	CHITTOOR	12.825828	0.144020	Urad	44	68	Kharif	65	19	32	508.0	32900	best y
1	CHITTOOR	12.820758	0.053614	Bajra	64	60	Kharif	64	50	60	514.0	45300	best y
2	CHITTOOR	12.805670	0.000000	Maize	16	91	Whole Year	64	50	60	514.0	422600	best y
3	CHITTOOR	12.805193	0.016096	Maize	25	98	Kharif	65	19	32	524.0	4100	best y
4	CHITTOOR	12.814065	0.000000	Moong(Green Gram)	74	85	Kharif	63	20	39	578.0	3700	best y

Figure 4.7: Dataset

4.7 Graphical representation of seasons

```
sns.countplot(x='Season',data=DF)
plt.xticks(rotation=90)
```

```
(array([0, 1, 2, 3]),
 [Text(0, 0, 'Kharif      '),
  Text(1, 0, 'Whole Year '),
  Text(2, 0, 'Rabi        '),
  Text(3, 0, 'Summer      ')])
```

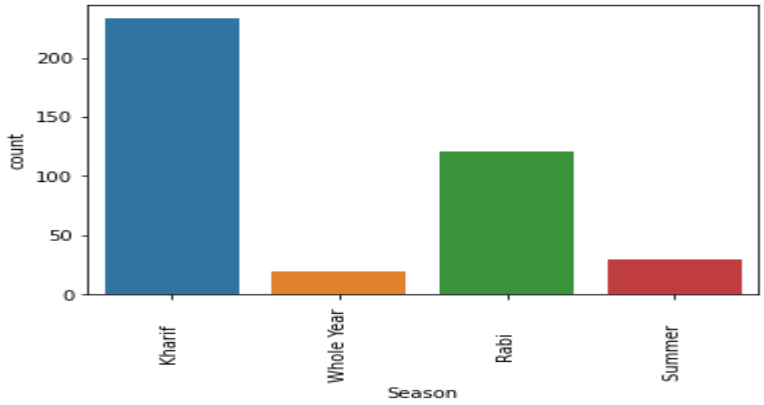


Figure 4.8: Diagram for Graphical (Bar Chart) representation of Seasons.

```
plt.figure(figsize = (6,6))
segment = DF['Season'].value_counts()
segment_label = DF['Season'].unique()
#color = ('LightPink', "LightBlue" , 'LightGreen','red','green','Gold')

plt.pie(segment,
        autopct = '%1.1f%%',
        labels = segment_label,
        shadow = True);
        #colors = color);
```

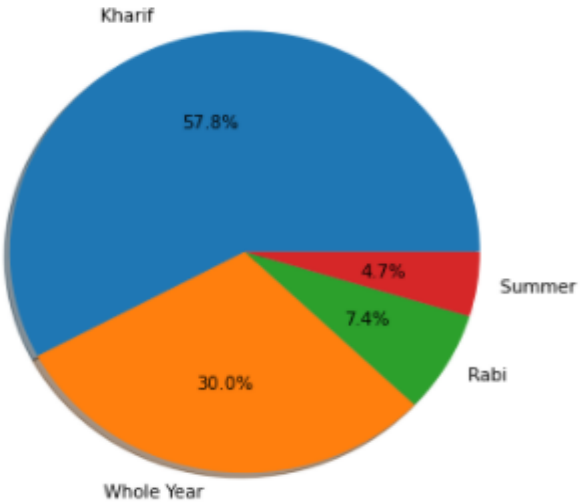


Figure 4.9: Diagram for Graphical (Pie Chart) representation of Seasons.

4.8 Graphical representation of crop

```
sns.countplot(x='Crop',data=DF)
plt.xticks(rotation=90)
```

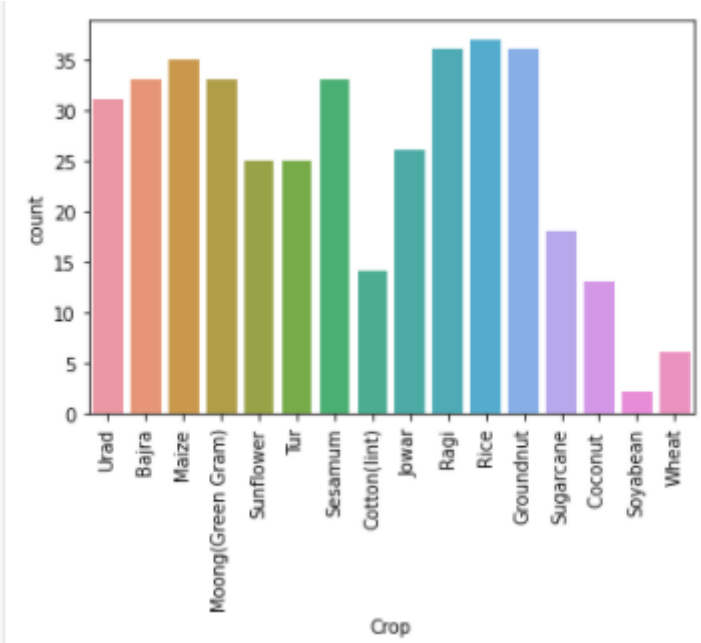


Figure 1: Diagram for Graphical (Bar Chart) representation of Crop.

```
plt.figure(figsize = (8,8))
segment = DF['Crop'].value_counts()
segment_label = DF['Crop'].unique()
#color = ('LightPink', "LightBlue" , 'LightGreen','red','green','Gold')

plt.pie(segment,
        autopct = '%1.1f%%',
        labels = segment_label,
        shadow = True,
        #colors = color
        );
```

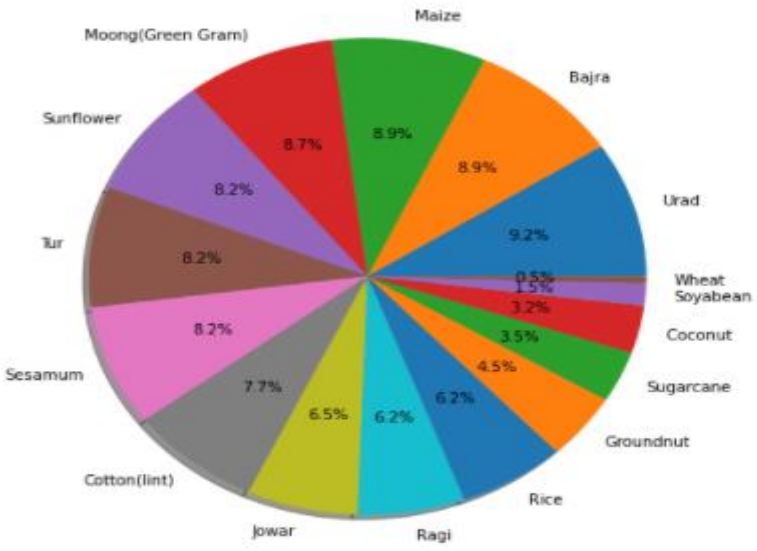


Figure 4.10: Diagram for Graphical (Pie Chart) representation of Crop.

4.9 Graphical representation of status of Production

```
sns.countplot(x='Status_of_yield',data=DF)
plt.xticks(rotation=90)
```

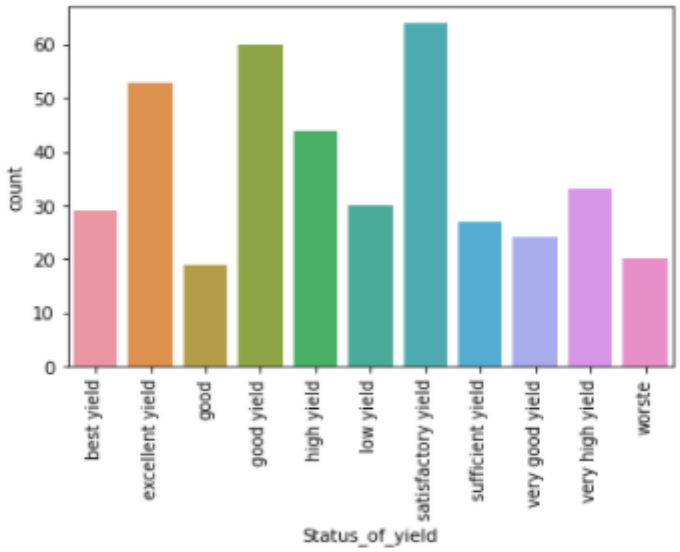


Figure 4.11: Diagram for Graphical (Bar Chart) representation of Status of Production.

```
plt.figure(figsize = (8,8))
segment = DF['Status_of_yield'].value_counts()
segment_label = DF['Status_of_yield'].unique()
#color = ('LightPink', "LightBlue" , 'LightGreen','red','green','Gold')

plt.pie(segment,
        autopct = '%1.1f%%',
        labels = segment_label,
        shadow = True,
        #colors = color
        );
```

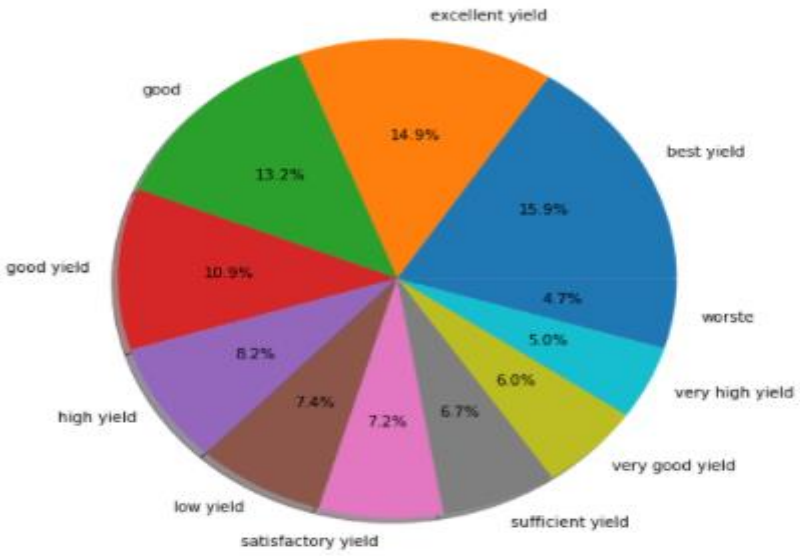


Figure 4.12: Diagram for Graphical (Pie Chart) representation of Status of Production.

```
import plotly.express as px
fig = px.pie(DFrame1, values='TRAINACCURACIES', names='ALGORITHMS')
fig.show()
```

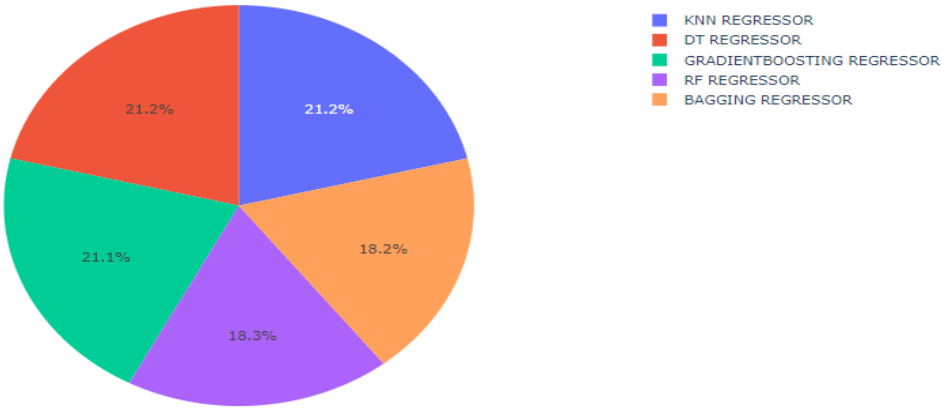


Figure 4.13: Diagram for Graphical (Pie Chart) representation of Algorithms.

4.10 Machine Learning Models Providing the Best Results

The selected studies used a diverse set of ML algorithms; their abbreviations are listed in Table A6 and A7. In the last two columns of Tables A1–A5, the ML algorithms used by each study, as well as those that produced the best results, are listed. As shown in Figure 5, the most common ML model producing the best results was, by far, Artificial Neural Networks (ANNs), which appeared in nearly 32.72 % of reviewed studies. ANN models, in particular, produced the best results in the majority of studies involving all sub-categories. ANNs were inspired by the biological neural networks that comprise human brains (Chen et al, 2019), and they enable learning through examples from representative data describing a physical phenomenon.

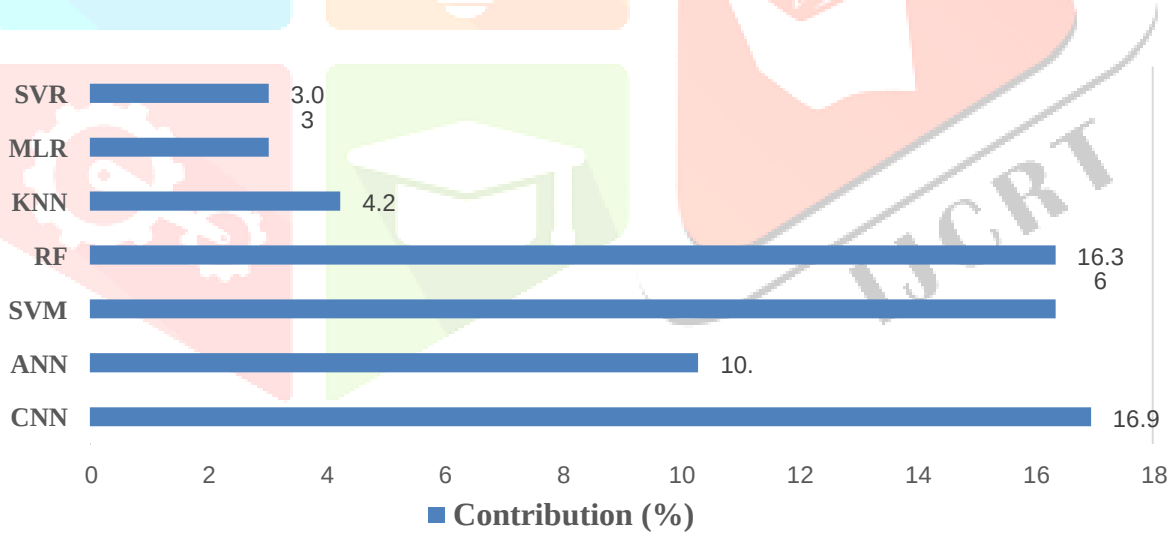


Figure 4.14: Contribution of different algorithms/models in review study

V CONCLUSION

The paper work introduces an efficient agriculture production prediction system using regression models. The system is scalable as it can be used to test on different crops. From the yield graphs the best time of sowing, plant growth and harvesting of plant can also be found out along with prediction for crops. The combination of regression algorithms like naive bayes, decision tree and random forest, bagging and gradient boosting are better performing than use of single regression model.

VI REFERENCES

1. A. Ahamed, N. Mahmood, N. Hossain, M. Kabir, K. Das, F. Rahman, R. Rahman, "Applying data mining techniques to predict annual yield of major crops and recommend planting different crops in different districts in Bangladesh", 16th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), pp. 1-6, 2015.
2. Aakunuri Manjula, Dr.G .Narsimha (2015), 'XCYPF: A Flexible and Extensible Framework for Agricultural Crop Yield Prediction', Conference on Intelligent Systems and Control (ISCO)
3. A.B. Mankar and M.S. Burange, "Data Mining-an evolutionary view of agriculture", ", International Journal of Application or Innovation in Engineering and Management, Vol 3, No 2 pp .102-105, March 2014
4. A.B.Mankar and M.S.Burange, "Data Mining – An Evolutionary View of Agriculture", International Journal of Application or Innovation in Engineering and Management, Vol. 3, No. 3, pp. 102-105, March 2014.
5. Bhuvana, Dr.C.Yamini (2015), 'Survey on Classification Algorithms in Data mining.' International Conference on Recent Advances in Engineering Science and Management
6. Cunningham, S. J., and Holmes, G. (1999). Developing innovative applications in agriculture using data mining. In the Proceedings of the Southeast Asia Regional Computer Confederation Conference,1999.
7. Data Mining – (Classifier /Classification- Function), https://gerardnico.com/wiki/data_mining/classification, last accessed on 6th October 2017.
8. Data Mining Techniques, https://www.ibm.com/developerworks/library/ba-data-miningtechniques/index.html, last accessed on 7th October 2017.
9. D. Ramesh and B. Vardhan, "Analysis of crop yield prediction using data mining techniques", International Journal of Research in Engineering and Technology, vol. 4, no. 1, pp. 47-473, 2015.
10. H. Patel and D. Patel, "A Brief Survey on Data Mining Techniques applied to Agriculture Data" International Journal of Computer Applications, Vol. 95, No. 9, pp. 6-8, June 2014.
11. Jharna Majumdar, Sneha Naraseeyappa and Shilpa Ankalaki, "Analysis of agriculture data using data mining techniques: application of big data", Jouranl of Big Day, Springer Open.
12. M.C.S.Geetha, "A Survey on Data Mining Techniques in Agriculture", International Journal of Innovative Research in Computer and communication Engineering, Vol. 3, No. 2, pp. 887892, February 2015.
13. Mucherino A, Papajorgji P, Pardalos PM: A survey of data mining techniques applied to agriculture. Oper Res. 2009, 9 (2): 121-140.
14. N.Gandhi and L.J. Armstrong, "Applying data mining techniques to predict yield of rice in Humid Subtropical Climatic Zone of India", Proceedings of the 10th INDIACom-2016, 3rd 2016 IEEE International Conference on Computing for Sustainable Global Development, New Delhi, India, 16th to 18th March 2016.
15. N. Gandhi and L. Armstrong, "Rice Crop Yield forecasting of Tropical Wet and Dry climatic zone of India using data mining techniques", IEEE International Conference on Advances in Computer Applications (ICACA), pp. 357-363, 2016.
16. N.Gandhi and L.J. Armstrong, "A review of the application of data mining techniques for decision making in agriculture", 2016 2nd InternationalConference on Contemporary Computing and Informatics (ic3i).
17. Onkar Kadam, Last updated Sept' 2017, 'Crop Data Analysis' https://www.kaggle.com/onkarkadam/crop-data-analysis
18. Rakesh Kumar, M.P. Singh, Prabhat Kumar and J.P. Singh (2015), 'Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique', International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM).
19. R.A.Medar and V.S.Rajpurohit, "A Survey on Data Mining Techniques for Crop Yield Prediction", International Journal of Advance Research in Computer Science and Management Studies, Vol.2, No. 9, pp. 59-64, September 2014.
20. R.Kalpana, N.Shanthi and S.Arumugam, "A Survey on Data Mining Techniques in Agriculture", International Journal of Advances in Computer Sciences and Technology, Vol. 3, No. 8, pp. 426-431, August 2014. [19] Anshal Savla, Parul Dhawan, Himtanaya Bhadada, Nivedita Israni, Alisha Mandholia , Sanya Bhardwaj (2015), 'Survey of classification

algorithms for formulating yield prediction accuracy in precision agriculture', Innovations in Information,Embedded and Communication systems (ICIIECS).

21. S.Pudumalar, E. Ramanujam, R. Harine Rajashreeń, C. Kavyań, T. Kiruthikań, J. Nishań, 'Crop Recommendation System for Precision Agriculture', 2016 IEEE Eighth International Conference on Advanced Computing (ICoAC).
22. Top 10 Data Mining Algorithms, https://www.kdnuggets.com/2015/05/top-10-data-miningalgorithms-explained.html, last accessed on 9th October 2017
23. Umid Kumar Dey, Abdullah Hasan Masud, Mohammed Nazim Uddin, "Rice yield prediction model using data mining", International Conference on Electrical, Computer and Communication Engineering (ECCE), February 16-18, 2017, Cox's Bazar, Bangladesh.
24. WEKA 3: Data Mining Software in Java, Machine Learning Group at the University of Waikato, Official Website: http://www.cs.waikato.ac.nz/ml/weka/index.html, accessed on 12nd October 2017.