Crop Prediction Using Machine Learning and Artificial Neural Network

Tanya Saraswat(✉)

Computer Science and Engineering, Amity School of Engineering and Technology, Amity University, Noida, India
saraswat.tanya24@gmail.com

Abstract. ***Agriculture Sector*** being most vital parts of every Nation. The production of a crop mostly relies on numerous features and result is relied on the end yield and on the selling rate of that crop. In today's world there is a growth of countless technologies which can predict the growth of a crop on a regular basis, what must be added in that type of soil to make it more productive on-the-basis of study of that region. ***Crop Prediction*** using ***Deep Learning*** methods is indeed an upcoming challenge in the field of Agriculture.

Deep Learning would increase the efficiency of the workforce, a huge amount of time would be spending in learning analytics, therefore increasing one's concentration leading to predicative analytics also there would be personalized learning, the dependency on others would slowly start to terminate. The main aim of this paper is to focus on crop prediction by using numerous algorithms of machine as well as ***deep learning***, and then to draw a comparison on the results and other performance measure of the different algorithms of ***Machine Learning*** and ***Deep Learning***.

Keywords: Agricultural Sector · Crop Prediction · Deep Learning · Machine Learning

1 Introduction

Crop Prediction is a vital part of the Indian economy as nearly all the sectors of the economy are related to the production in the agricultural sector and what sorts of profit happens in the Agricultural Sector. The Crop Prediction mainly hangs upon a few weather situations and circumstances such as:

1. Rainfall Prediction in the near future
2. Temperature variations
   a) Next Heat wave prediction
   b) Next Cold wave prediction

The Crop Prediction is also dependent upon numerous other factors like:

1. Nitrogen Concentration in soil
2. Potassium Dilution in the soil

R. Manza et al. (Eds.): ACVAIT 2022, AISR 176, pp. 561–568, 2023.
https://doi.org/10.2991/978-94-6463-196-8_43

3. Phosphorus present in the soil
4. Ph of the soil

Precise evidence in relation to the historical prediction of the crop is imperative for taking in effective conclusion that are in relation with the hazards that the agriculture sector may face in the near-future and how to pass through those risks and to make the approximate forecasting and prediction of it beforehand to prevent the losses. In the current work, Crop Prediction is being made by using different algorithms of Machine Learning as well as Deep Learning, both of which are an important segment of prediction [1]. In the current work, the predictions have been made on basis of some of the elements which may or may not lead to the increment in the agricultural sector of the economy of India. In India, largely and mainly depends upon the agricultural sector hence forth, there is a huge need for examining the element related to crops which may result to the boon in the primary sector [1].

The key objectives which have been used are divided into five parts, data being pre-processed, data visualization, training and testing of the data, training of the model and then the evaluation of the model with effect of numerous algorithms.

The main objective of the current work is to draw the prediction of the given crops by using random forest, decision tree classifier, Gaussian Naïve Bayes, Gradient Booster, Logistic Regression and Artificial Neural Network to find out which of the models predict the stroke with the highest accuracy. Current work has been divided into following parts, Section 2 tells about the Literature Review, Sect. 3 deals with methodology, which is adopted, Sect. 4 presents the result followed by Sect. 5 which provides the inference of the work presented [15].

2 Literature Review

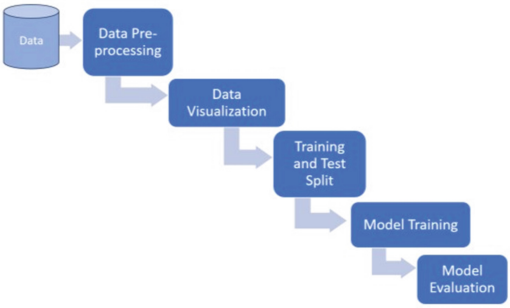
Crop Prediction deals with several complicated attributes which include the genotype of the crop, environmental changes and conditions, and the interaction of these genomes and ecology. These traits widely describe the properties which are going to be used in this Research Paper [12, 13].

For producing crop on a worldwide scale, crop yield prediction is crucial. In this Research paper, drawing a crucial analysis using the different machine learning algorithms, Ensemble Learning as well as Artificial Neural Network [14].

In this Research Paper, interdependency of the various weather conditions on the Crop is being reviewed and researched upon.

3 Methodology

In the current work, the methodology which is used has been divided into five segments or parts named as, Data being Pre-processed, Data Visualization, Training and testing the data, training of the model and then the evaluation of the model on basis of numerous algorithms [2].



Flow-diagram of Methodological structure used

Details about the dataset used in this work is mentioned below:

a) *Dataset Description* The dataset used in the current work has been taken up from Kaggle. The dataset provides the information which deals with the prediction of the crop.

The dataset consists of seven elements on basis of which the crop prediction is dependent upon, they are: - N (Nitrogen), P (Phosphorus), K (Potassium), Temp (Temperature), Humidity, Ph and Rainfall. The dataset consists of 2200 entities wherein there are 22 crops and seven other elements on which the crop prediction depends upon [3].

There are 100 entries for every type of crop. The size of the Dataset suitable for performing various performance measures such as accuracy and precision, as these performance measures require a dataset with heavy data.

The value count of all the respective crops is shown beneath:

The dataset comprises of these different aspects:

```
blackgram      100
pigeonpeas     100
mango          100
rice           100
chickpea       100
watermelon     100
lentil         100
coconut        100
mothbeans      100
coffee         100
orange         100
muskmelon      100
mungbean       100
pomegranate    100
grapes         100
maize          100
kidneybeans    100
banana         100
cotton         100
apple          100
papaya         100
jute           100
Name: Crop, dtype: int64
```

Fig. 1. Number of Crops in the dataset

Nitrogen: This aspect or property is a numerical attribute, which tells us about how much nitrogen is present in the soil and Nitrogen plays a vital role in the growth and development of the plant as Nitrogen is foremost constituent element which helps in letting the plants generate sugar by the help of sunlight from the presence of water as well as Carbon Dioxide.

Phosphorus: This aspect or property is a numerical attribute, which is used in storing and then transferring of energy in the crops. Henceforth, it is the topmost necessity in crop prediction.

Potassium: This aspect is used to determine the water holding capacity in the plants. The need of phosphorus in plants is to hold-down the water, due to the absorption rate.

Temperature: Temperature is an important aspect which is used to determine at what rate the evaporation will take place. For example, in the daytime when the temperature is more, the evaporation rate is more henceforth, crops whose surface area of leaf is more will need more water as the evaporation rate will be higher, and similarly at nighttime, when the sunsets, the evaporation rate decreases, decreasing the need of water in the soil. It is therefore a key factor in determining the prediction of crop.

Humidity: This is again a numerical aspect which find out the moisture present in the environment, and what is the result of it on the crop.

Ph: The Ph also defines a lot about the crop therefore, it is also an important factor in diffing the type of crop being grown.

Rainfall: Rain consists of many useful nutrients which are beneficial for the crop. It is also observed that the germination process of the crop tends to begin after the monsoon season.

Also, if the rainfall tends to happen for a longer time, then it may result in a havoc for the crops as they can get rot. Farming depends upon the period wherein the rain is occurring. It plays a significant role in deciding as well as predicting the crop.

The amount of rainfall also has a huge positive impact on the rising-up of the water table, as the water table rises, the plants which needs more water, can get water from the soil itself. Rainfall also increases the soils fertility by adding extra nutrients into it, thus enhancing the quality of the soil [4].

b) *Data Preprocessing* Later when the dataset has been collected from several resources, it's a necessity to pre-process the available Dataset after collection. Pre-processing of the available data could be completed at several stages. This starts with the reading and understanding the data which has been gathered.

Next step is the cleaning of the data which is being used in the current work [1, 4]. For example, if the dataset is having a huge amount of NAN values, then those values or the entire column must be dropped down beforehand itself, or it may be filled with the maximum or random values as to get better accuracy [5].

It is foreseen that, a data which consists of larger amount of NAN values, then that doesn't result in higher accuracy as data-cleaning is being performed due to the faults present in the data. If data preprocessing is being performed using python language, then some pre-defined libraries must be imported to the server in order to perform pre-processing [3, 4].

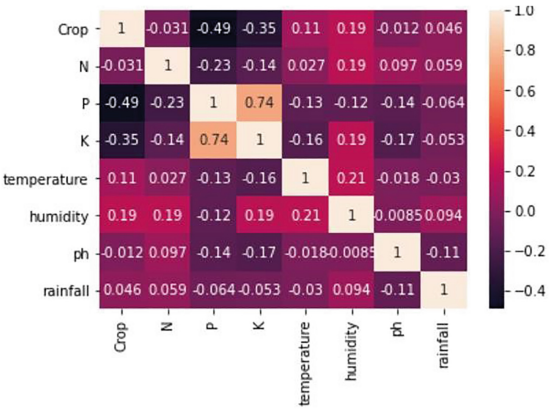


Fig. 2. Heatmap

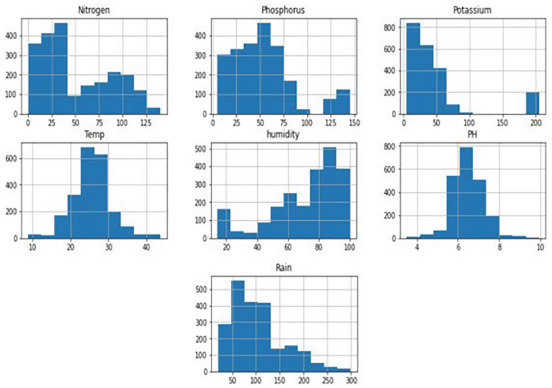


Fig. 3. Histogram of numerical data

c) *Train – Test Split* In this step of the methodology, the data is now categorized into two different segments known as the:

a) Training Data
b) Test Set

Training and test splits main motive is to evaluate and calculate the execution of Machine Learning as well as Deep Learning Algorithms. Training and Test Split technique is beneficial for Regression as well as the Classification problems of Machine Learning.

Technique coming under Supervised Learning [4, 5].

I. Training Set: As the term itself suggest that the training of the entire data is to be done on this particular data.

II. Testing Set: The word is self-explanatory; this work means that the testing is being done on which parameter for providing the forecast or prediction.

The cause behindhand using the train and test split is to accomplish the predictions and forecasting on new data rather than on the data which had by now been used for training purposes. In the current work, the data on which train and split is being performed had been fragmented into proportion of 80:20 which means the training data occupies 80% whereas the testing data occupies 20% [10]. For the fragmentation of the given data, the method that was used was, from 'sklearn.model_selection', the test and train set was imported ant the this was used for training and testing of the dataset 'train_test_split'.

In the subset of Machine learning, Supervised Learning, model training is meant to be feeding the machinery with the different algorithms present along with the training as well as the testing data available so that the targeted value could acquire from it [7].

The main motive of Training of the model is generation of trained model as in to simplify well known to new known to not-known data. The trained-fitted model estimated by the help of these new known data from the beforehand, handheld data as to evaluate the accuracy provided by training of the model.

In the current work, the classifiers used are Random Forest, Decision Tree, Gaussian Naïve Bayes, Gradient Booster, Support Vector Machine, Logistic Regression and Artificial Neural Network (ANN) [6].

4 Results

Machine learning and deep learning are a very useful approach for working with prediction as well as forecasting.

The limitation of these algorithms is that, in dataset the required, data must have enormously a huge number of entries' else, the performance measure will be having lower performance.

In the current work, the methodology adopted is for the prediction of the most appropriate crop with the help of identification of several aspects of element that are essential for the best growth as well as the development of the crop such as the soil containing adequate amount of nutrients or not, also specifications which are related to the atmosphere like, humidity as well as temperature.

This piece of work hereby displays the capability of the algorithms used in the above work for the forecasting as well as the prediction of crops in different sections of India, where in the soil requirement and the atmospheric requirements are fulfilled [11].

5 Conclusion and Future Work

Prediction of crop is a necessity as it is the daily need of everyone. No human being can live a life without being feed on a crop. Also, India's economy is directly indirectly dependent upon Crop thereafter, it is not only a necessity but also the essential requirement of living [8]. In the rising demand of crop prediction, to get precaution and cautious about the near-future risks, the prediction of crop had been done on basis of seven other

Table 1. Comparison B/W Classifiers (Accuracy)

MODELS APPLIED	Accuracy (in %)
Random forest classifier	99.3181818
Decision tree classifier	98.8636364
Gaussian naive bayes	99.5454545
Gradient booster classifier	99.4886364
Support vector machine	96.1363636
Logistic regression	94.5454545
Deep learning ANN	99.3574803

Table 2. Distribution Table verified by different algorithms

	Range						
Crop Name	N	P	K	Temperature	Humidity	pH	Rainfall
Apple	0-40	120-145	195-205	21.04-24.00	90.03-94.93	5.52-6.50	100.12-124.99
Banana	80-120	70-95	45-55	25.02-29.91	75.04-84.98	5.51-6.50	090.11-119.85
Blackgram	20-60	55-80	15-25	25.10-34.95	60.07-69.97	6.51-7.78	060.42-074.92
Chickpea	20-60	55-80	75-85	17.03-21.00	14.26-19.97	5.99-8.87	065.12-094.79
Coconut	0-40	5-30	25-35	25.01-29.87	90.02-99.99	5.51-6.48	131.10-225.64
Coffee	80-120	15-40	25-35	23.06-27.93	50.05-69.95	6.03-7.50	115.16-199.48
Cotton	100-140	35-60	15-25	22.01-26.00	75.01-84.88	5.81-8.00	060.66-099.94
Grapes	0-40	120-145	195-205	08.83-41.95	80.02-83.99	5.52-6.50	065.02-074.92
Jute	60-100	35-60	35-45	23.10-26.99	70.89-89.90	6.01-7.49	150.24-199.84
Kidneybeans	0-40	55-80	15-25	15.34-24.93	18.10-24.97	5.51-6.00	060.28-149.75
Lentil	0-40	55-80	15-25	18.07-29.95	60.10-69.93	5.92-7.85	035.04-054.94
Maize	60-100	35-60	15-25	18.05-26.55	55.29-74.83	5.52-7.00	060.66-109.76
Mango	0-40	15-40	25-35	27.01-36.00	45.03-54.97	4.51-6.97	089.30-100.82
Mothbeans	0-40	35-60	15-25	24.02-32.00	40.01-64.96	3.51-9.94	030.93-074.45
Mungbean	0-40	35-60	15-25	27.02-29.92	80.04-90.00	6.22-7.20	036.13-059.88
Muskmelon	80-120	5-30	45-55	27.03-29.95	90.02-94.97	6.01-6.79	020.22-029.87
Orange	0-40	5-30	5-15	10.02-34.91	90.01-94.97	6.02-8.00	100.18-119.70
Papaya	31-70	46-70	45-55	23.02-43.68	90.04-94.95	6.51-7.00	040.36-248.86
Pigeonpeas	0-40	55-80	15-25	18.32-36.98	30.41-69.70	4.55-7.45	090.06-198.83
Pomegranate	0-40	5-30	35-45	18.08-24.97	85.13-95.00	5.57-7.20	102.52-112.48
Rice	60-99	35-60	35-45	20.05-26.93	80.13-84.97	5.01-7.87	182.57-298.57
Watermelon	80-120	5-30	45-55	24.05-26.99	80.03-89.99	6.01-6.96	040.13-059.73

parameters. The prediction had been done using seven algorithms. The training and testing split data taken for this purpose was in the proportion of 80:20. Highest accuracy was approximately 99.54% which was given by Gaussian Naïve Bayes Algorithm and second highest accuracy was about 99.35% and was given by Artificial Neural Network [9]. Prediction made by these algorithms are particular in every aspect and can be used to perform predictions in-actuality.

References

1. Kaggle.com. 2021. *Crop Prediction Dataset*. [online]

2. Ashwani kumar Kushwaha, Swetabhattachrya "crop yield prediction using agro algorithm in hatoop"
3. Pavan Patil, Virendra Panpatil, Prof. Shrikant Kokate "Crop Prediction System using Machine Learning Algorithms"
4. Girish L, Gangadhar S, Bharath T R, Balaji K S, Abhishek K
5. T "Crop Yield and Rainfall Prediction in Tumakuru District using Machine Learning".
6. Manjula E, Djodiltachoumy S (2017) A model for prediction of crop yield. Int J Comput Intell Inf 6(4)
7. Clarkson, D. T. & Warner, A. (1979). Relationships between root temperature and the transport of ammonium and nitrate ions by Italian and perennial ryegrass (*Lolium multiflorum*and *Lolium perenne*). Plant Physiology 64, 557– 561.
8. Aggarwal Sachin (2001). Application of Neural Network to Forecast Air Quality Index. Thesis submitted in partial fulfillment of requirements for a degree in Bachelor of Technology, April 2001.
9. Everingham, Y. L., R. C. Muchow, R. C. Stone, and D. H. Coomans, 2003. Using southern oscillation index phases to forecast sugarcane yields: a case study for Northeastern Australia. International Journal of Climatology 23(10): 12111218.
10. D.L. Ehret et al, Neural network modeling of greenhouse tomato yield, growth and water use from automated crop monitoring data, 2011.
11. Rice Research and Development Institute (RRDI), *Department of Agriculture-Sri Lanka.Rice Cultivation*, Rice Research and Development Institute (RRDI), Ibbagamuwa, Sri Lanka, 2020.
12. Y. Masutomi, K. Takahashi, H. Harasawa, and Y. Matsuoka, "Impact assessment of climate change on rice production in Asia in comprehensive consideration of process/parameter uncertainty in general circulation models," *Agriculture, Ecosystems & Environment*.
13. S. Khaki and L. Wang, "Crop yield prediction using deep neural networks", *Frontiers in plant science*.
14. T. Senthil Kumar, "Data Mining Based Marketing Decision Support System Using Hybrid Machine Learning.
15. Algorithm", *Journal of Artificial Intelligence*, 2020.
16. V. Pandith, H. Kour, S. Singh, J. Manhas and V. Sharma, "Performance Evaluation of Machine Learning Techniques for Mustard Crop Yield Prediction from Soil Analysis", *Journal of Scientific Research*, 2020.
17. S. D. Kumar, S. Esakkirajan, S. Bama and B. Keerthiveena, "A microcontroller based machine vision approach for tomato grading and sorting using SVM classifier", *Microprocessors and Microsystems*, 2020.