IJDCST @ May-June-2021, Issue- V-7, I-1, SW-24
ISSN-2320-7884 (Online)
ISSN-2321-0257 (Print)

Smart Farming using Machine Learning

Roshan Mallipeddi, Swetha Prathipati, Narasimha Yaddanapudi, Swapna Tondapu, Swapna Donepudi,
Dept. of CSE, PVPSIT – Vijayawada, Andhra Pradesh – 520 007.

***Abstract* — Farming is one of the significant distinct advantage and a significant income delivering area in India. India's agribusiness area is in emergency for almost twenty years now. Various seasons, markets, and organic examples impact crop creation, yet changes in these examples bring about a fantastic misfortune to ranchers. Self-destructive cases are filling in numbers throughout the long term. This is a result of the absence of appropriate information and the unremarkable strategies embraced by ranchers in their cultivating. In this manner, Millet Harvest Yield Forecast is a significant rural issue. This issue can be abstained from by embracing keen cultivating strategies for example fusing innovation in every day cultivating. The undertaking for the most part centers around determining valuable experiences on crop yield forecast. The factual, rural dataset is attempted for trial investigation. The information is pre-handled and grouped into preparing and testing information. Different classification algorithms are used for calculating the millet crop yield prediction by taking various input fields like soil, min temp, max temp, humidity, rainfall, etc. With the help of all these features, smart farming can be achieved.**

Keywords— Machine Learning, Crop Yield Prediction, Millet Crop Yield, Random Forest, Support Vector Machine, Logistic Regression, Neural Networks.

I. INTRODUCTION

Agriculture is an important sector for the Indian economy and the human future. It is first and foremost work which is essential for life. It also contributes a large portion of employment. As time passes the need for production has been increasing exponentially. To produce in mass quantity people are using technology in the wrong way. New kinds of hybrid varieties are produced day by day. However, these varieties do not provide the essential contents as naturally produced crops. These unnatural techniques spoil the soil. It all leads to further environmental harm. Most of these Unnatural techniques are used to avoid losses. But when the producers of these crops know the accurate information on the crop yield it minimizes the loss. To achieve this project is made. Using past information on weather, temperature, and several other factors the information is given. Data mining is a machine-learning tool that is used to view data in all possible ways and analyze it. After analyzing the data, it is used to predict the future. Purposes. It can be used in several fields. These patterns provide information about crops. The aim of the project is to results to increase the yield and profit for producers. The proposed system concentrates on yield prediction. The dataset is taken on agriculture statistics. This dataset is used as an experimental basis. After the data processing, it is divided into training and testing.

II. LITERATURE REVIEW

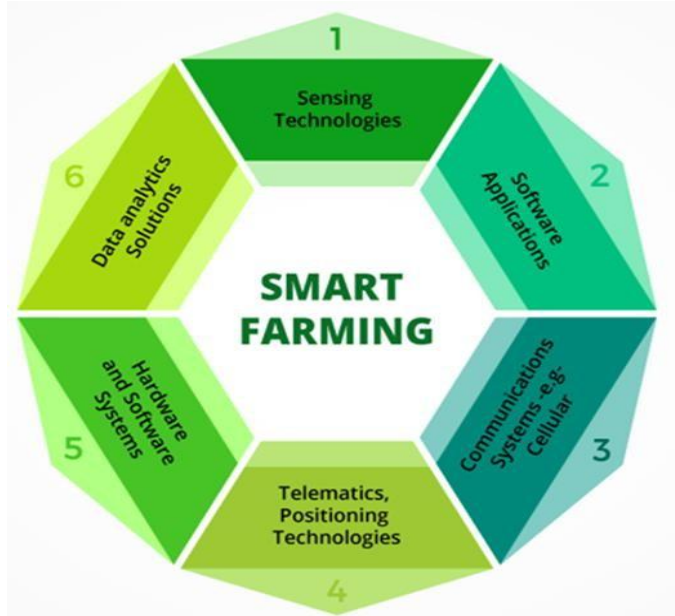
A. Precision Agriculture

Precision agriculture makes use of a range of technologies that include GPS services, sensors, and big data to optimize crop yields. Rather than replace farmer expertise and gut feeling, ICT-based decision-support systems, backed up by real-time data, can additionally provide information concerning all aspects of farming at a level of granularity not previously possible. This enables better decisions to be made, resulting in less waste and maximum efficiency in operations. The disciplines and skills now required for agriculture include computer-based imaging, GPS technology, science-based solutions, climate forecasting, technological solutions, environmental controls, and more. Precision agriculture is sometimes known as 'smart farming', an umbrella term for easier comparison with other M2 M-based implementations such as smart metering, smart cities, and so on. It is based on sensor technologies whose use is well established in other industries, e.g. Environmental monitoring for pollutants, eHealth monitoring in patients, buildings management for farm soil monitoring, and so on. For all M2M implementations, IT systems gather, collate, analyze the data and present it in such a way as to initiate an appropriate response to the information received. For farmers and growers, a wide variety of information regarding soil and crop behavior, animal behavior, machine status, storage tank status emanating from remote sites is presented for action by the farmer.

ISSN-2320-7884 (Online)
ISSN-2321-0257 (Print)

B. Smart Farming Eco System

The complexity of smart farming is also reflected into the ecosystem of players. They can be classified in the following way: Technology providers – these include providers of wireless connectivity, sensors, M2M solutions, decision support systems at no mapping applications. Providers of agricultural equipment and machinery, farm buildings, as well as providers of specialist products (e.g. seeds, feeds) and expertise in crop management. Customers: farmers, farming associations. Influencers – those that set prices, influence market into which farmers and growers to sell their products. The end users of precision farming solutions include not only the growers but also farm managers, users of back-office IT systems. Not to be forgotten is the role of the veterinary in understanding animal health. Also to be considered are farmers\' co-operatives, which can help smaller farmers with advice and funding. The cost of smart farming is still high for any but the largest farms. Farm offices now collect vast quantities of information from crop yields, soil mapping, fertilizer applications, weather data, machinery, and animal health; these are all factors that influence farming such as soils, nutrition, and weather. Data is the fundamental building block of smart farming, whether the data comes from a soil sample or a satellite correction signal. For example, data points collected can highlight both spatial and temporal variability within a field. Many factors can contribute to this variability; understanding the effect each factor has can only be measured and managed using statistical analysis of the data. Everyday farming applications are starting to move into the cloud, with the aim of delivering benefits in terms of data access, synchronization, storage, and even cost to the farmer. The rising use of smartphones and tablets on farms means that apps can be used to cache data offline until it can be synchronized; data need no longer be tied to a single computer in a single location.



III. Proposed Approach

This scenario mainly concentrates on crop forecasting. These factors help the farmers to cultivate the best food crops and raise the right animals in accordance with environmental components. Also, the farmers can adapt to climate changes to some degree by shifting planting dates, choosing varieties with different growth duration, or changing crop rotations. For the experimental analysis, the statistical numeric data related to agriculture is undertaken. Whereas, clustering-based techniques and supervised algorithms are utilized for managing the collected statistical data. Additionally, suitable classification methods like Random Forest (RF), Support Vector Machine (SVM), Logistic Regression (LR), Neural Networks (NN) are employed for better classification outcomes.

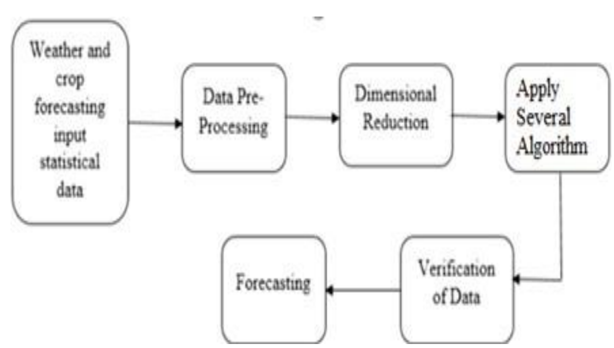


Fig. Forecasting Block Diagram

A. Dataset & Pre - processing

For the experimental purpose, the statistical information is collected from kaggle. We have taken a dataset consisting of historical data for millets. The dataset have about 3900 instances. It consists of various attributes such as moisture,

IJDCST @ May-June-2021, Issue- V-7, I-1, SW-24
ISSN-2320-7884 (Online)
ISSN-2321-0257 (Print)

rainfall, temperature, average humidity, mean, min & max temperatures, different attributes related to soil such as alkaline, sandy, chalky, clay & the output attribute millet yield. The millet yield comprises values ranging between 0 - 4 which define the poor, below average, average, good & excellent yields.

	Moisture	rainfall	Average H	Mean Tem	max Temp	Min temp	alkaline	sandy	chalky	clay	millet yielc
0	12.80168	0.012361	57	62	71	52	1	0	0	0	
1	12.85165	0.004172	57	58	73	43	1	0	0	0	
2	12.77677	0	56	58	69	46	1	0	0	0	
3	12.942	0.031747	62	56	70	43	0	1	0	0	
4	12.98465	0	65	56	70	42	0	0	1	0	
5	12.96447	0.027191	65	58	70	46	1	0	0	0	
6	12.738	0.026821	61	56	70	42	0	1	0	0	
7	12.81938	0.010284	58	57	72	42	0	1	0	0	
8	12.88391	0.020465	63	60	76	45	0	0	0	1	
9	12.78451	0.060054	62	59	71	47	1	0	0	0	
10	12.96881	0.084119	56	58	69	46	1	0	0	0	
11	12.78436	0	63	56	70	42	0	0	0	1	
12	12.94459	0	67	58	72	43	0	1	0	0	
13	12.92528	0.124479	58	60	75	46	0	1	0	0	
14	12.82107	0.074505	59	58	68	49	1	0	0	0	
15	12.93922	0.098584	53	60	68	53	0	0	0	1	
16	12.81587	0.222846	62	60	69	50	0	0	0	1	
17	13.05681	0.128555	59	56	67	45	0	0	1	0	
18	12.89796	0.113419	58	54	68	40	0	0	0	1	
19	13.03464	0.080064	58	53	68	38	0	0	0	1	
20	13.02395	0.077676	66	55	63	47	0	0	0	1	

In data pre – processing, we import the required packages, read the dataset and drop the unnecessary columns and check whether null values are present in our data or not. The isnull() method is used to check for null values. The sum() method is used for getting the shortest output result. We categorize our data into two parts 80% for training & 20% for testing.

```
dt.isnull().sum()
```

```
Moisture            0
rainfall            0
Average Humidity    0
Mean Temp           0
max Temp            0
Min temp            0
alkaline            0
sandy               0
chalky              0
clay                0
millet yield        0
dtype: int64
```

B. *Feature Selection*

It is the process of reducing the raw data into reasonable groups for processing it. Starting with an initial set of raw data, it builds up derived features which results in an informative and non – redundant data. As the statistical agriculture data is redundant and too large to be processed, it is first transformed into a reduced or minimal set of features.

C. *Classification*

Classification is the process of categorizing a data object into categories into classes based upon features / attributes associated with that data object. Classification uses a classifier, an algorithm that processes the attributes of each data object and outputs a class based upon the information. In this project, we use Random Forest (RF), Support Vector Machine (SVM), Logistic Regression (LR), Neural Networks (NN) as classifiers.

RANDOM FOREST:

Random Forest is a binary tree-based machine-learning methodology. We use this algorithm to predict yields of varied crops. RF develops many decision trees based on a random selection of data and variables. More trees result in a more robust prediction. Random forest handles the missing values and handles the accuracy for it. It handles the dataset with higher dimensionality.

Algorithm:

Random Forest Algorithm (RF) is also known as Random Forest Classifier. Coming up next are the means to comprehend RFA.

- Step – 1: Start with defining arbitrary examples from the given dataset.
- Step – 2: Then, this computation will foster a decision tree for every model. By then, it will get the normal outcome from every decision tree.
- Step – 3: In this step, voting will be performed for every predicted result.
- Step – 4: This is final step in RFA, select the most voted prediction as the final prediction result.

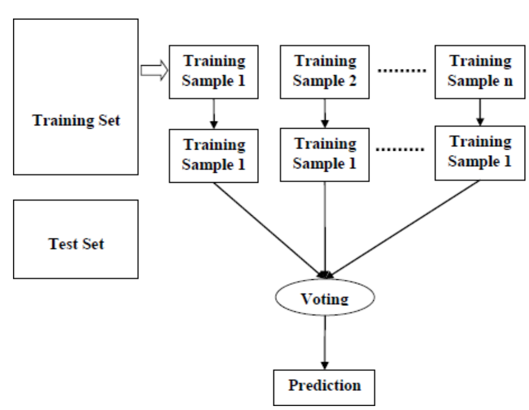


Fig. Working scenario of Random Forest Algorithm.

IJDCST @ May-June-2021, Issue- V-7, I-1, SW-24
ISSN-2320-7884 (Online)
ISSN-2321-0257 (Print)

This estimation runs capably on immense information bases and it has higher gathering exactness. Exact data about the history of harvest yield is something critical for settling on choices identified with farming danger the executives. In this way, the paper proposes a plan that forsees the yield of the harvest. The farmer will check the yield of the harvest according to the section of land, before developing onto field.

SUPPORT VECTOR MACHINE:

Support Vector Machine is a classification technique. It is a model that best split the different features. Its main objective is to figure out the perfect hyperplane which distinctly classifies the data points. The distance between the data points (support vector) and the hyperplane are as far as possible. One challenge with Support vector machine algorithm is that if the features or dimensions increase it is hard to visualize.

LOGISTIC REGRESSION:

Logistic regression is a process of modeling the probability of a discrete outcome given an input variable. The most common logistic regression models a binary outcome; something that can take two values such as true/false, yes/no, and so on. Multinomial logistic regression can model scenarios where there are more than two possible discrete outcomes. Logistic regression is a useful analysis method for classification problems, where you are trying to determine if a new sample fits best into a category.

NEURAL NETWORKS:

A neural network is a series of algorithms that endeavors to recognize underlying relationships in a set of data through a process that mimics the way the human brain operates. In this sense, neural networks refer to systems of neurons, either organic or artificial in nature.

A multilayer perceptron is a neural network connecting multiple layers in a directed graph, which means that the signal path through the nodes only goes one way. Each node, apart from the input nodes, has a nonlinear activation function. An MLP uses backpropagation as a supervised learning technique. Since there are multiple layers of neurons, MLP is a deep learning technique.

IV. RESULTS

The end – product of the task is to anticipate the crop yield. We arrange the crop – yield as poor yield, below average yield, average yield, good yield & excellent yield. The smart farming crop yield prediction is to predict the crop yield and use the predictions in developing better quantitative and qualitative crops.

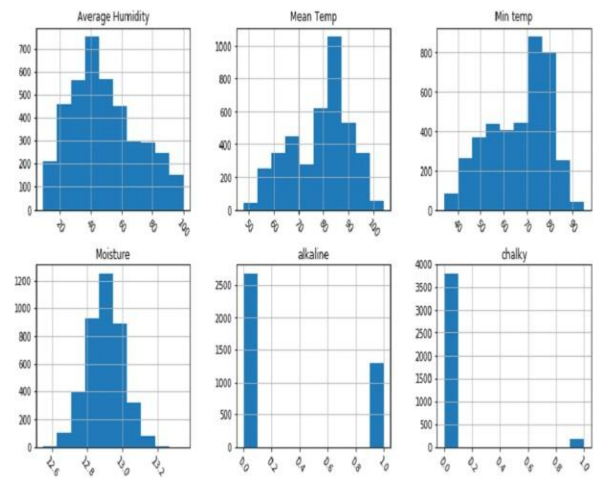
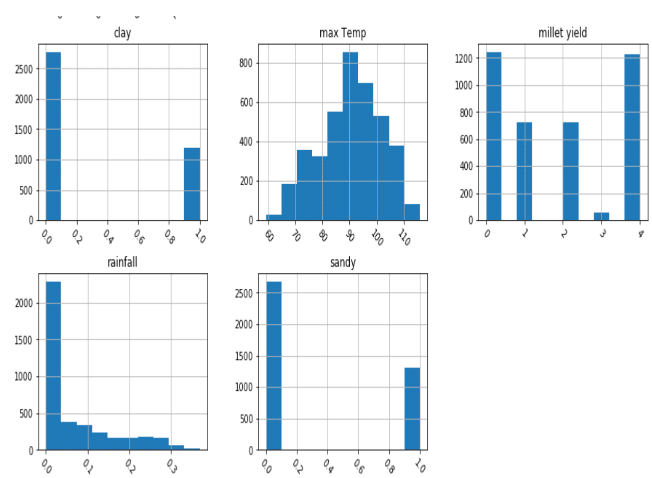
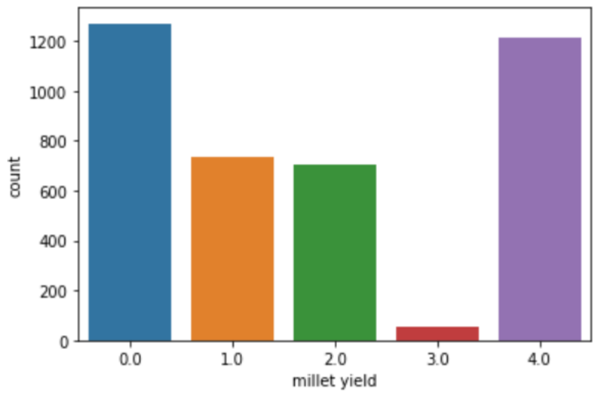


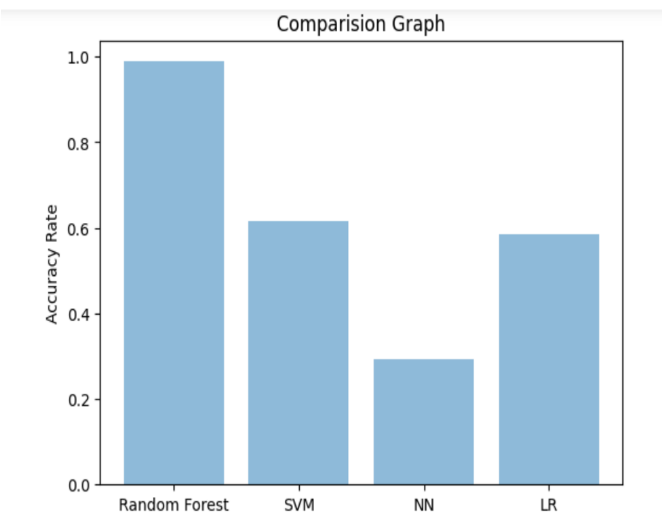
Fig. shows the histogram charts of various fields like Average Humidity, Mean Temp, Min Temp, Moisture, Alkaline, Chalky. All these fields are taken as raw input data sets. Each histogram graph has 10 bins. The input raw datasets also consist of various fields like Dry, Max Temp, Millet Yield, Rainfall and, Sandy as shown below.



In data visualization, we visualize the millet yield availability first using seaborn for better understanding.



The comparision graph comprising the accuracies obtained by the algorithms is implemented using matplotlib and is as follows:



The Random Forest Algorithm provides us the high accuracy of 98.8%. We dump this Random Forest model into a .pkl file which can be predominantly used for prediction. The prediction of the crop yield can be done in a web enhancement. The web enhancement is implemented using Python's Flask framework.

Flask framework is used for building web application using Python by using the Machine Learning models.

V. CONCLUSION

Agriculture has always been the most important sector for survival. There are a lot of difficulties faced by our farmers due to various unpredictable reasons. Consequently, as engineers, we need to team up with farmers and provide them a solution to improve the quality & quantity of crops. Our project is the first step towards it. Prediction can help us make strategic decisions in crop production. With Machine Learning, we get insights into crop life which can be very beneficial. We can construe that when the alkaline is least (0.0) the millet yield is greatest (2500). And when the chalky content is least (0.0) the millet yield is greatest (3500). And when chalky content is extreme then millet yield is less. Then, the dampness content is normal then we acquire most extreme yield for millet.

VI. FUTURE WORK

As the smart farming methodologies increase, there would be a vast requirement for newer technologies to be implemented. The project which is now a web – based application can be made into an android app where farmers can be educated and informed about their crop yield and provide the measures for good crop yield. Once a prediction is done we can improve on the automation process where the farmers can remotely control the field using a mobile app.

VII. REFERENCES

[1] D. R. Legates, R. Mahmood, D. F. Levia, T. L. DeLiberty, S. M. Quiring, C. Houser, and F. E. Nelson, "Soil moisture: A central and unifying theme in physical geography," Progress in Physical Geography, vol. 35, no. 1, pp. 65–86, 2011.

[2] Y. H. Kerr, P. Waldteufel, J.-P. Wigneron, J. Martinuzzi, J. Font, and M. Berger, "Soil moisture retrieval from space: The soil moisture and ocean salinity (smos) mission," IEEE transactions on Geoscience and remote sensing, vol. 39, no. 8, pp. 1729–1735, 2001.

[3] E. G. Njoku, T. J. Jackson, V. Lakshmi, T. K. Chan, and S. V. Nghiem, "Soil moisture retrieval from amsr-e," IEEE transactions on Geoscience and remote sensing, vol. 41, no. 2, pp. 215–229, 2003.

[4] R. H. Reichle, R. D. Koster, J. Dong, and A. A. Berg, "Global soil moisture from satellite observations, land surface models, and ground data: Implications for data assimilation, " Journal of Hydrometeorology, vol. 5, no. 3, pp. 430–442, 2004.

[5] S. Lambot, E. C. Slob, I. van den Bosch, B. Stockbroeckx, and M. Vanclooster, "Modeling of ground-penetrating radar for accurate characterization of subsurface electric properties, " IEEE Transactions on Geoscience and Remote Sensing, vol. 42, no. 11, pp. 2555–2568, 2004.

[6] M. S. Dawson, A. K. Fung, and M. T. Manry, "A robust statisticalbased estimator for soil moisture retrieval from radar measurements," IEEE transactions on geoscience and remote sensing, vol. 35, no. 1, pp. 57–67, 1997.

[7] G. Satalino, F. Mattia, M. W. Davidson, T. Le Toan, G. Pasquariello, and M. Borgeaud, "On current limits of soil moisture retrieval from ers-sar data," IEEE Transactions on Geoscience and Remote Sensing, vol. 40, no. 11, pp. 2438–2447, 2002.