JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

CROP YIELD PREDICTION USING MACHINE LEARNING

Dr. P. Punitha Ponmalar
Associate Professor, PG & Research Department of Computer ScienceSri Meenakshi Government Arts College for women(Autonomous) Madurai, India.

R. Siva Ranjani
Student- M.Sc., PG & Research Department of Computer Science Sri Meenakshi Government Arts College for women(Autonomous)Madurai, India.

ABSTRACT In this research paper, the machine learning models to predict the best crop and best fertilizers to be used based on the soil quality and location. This model explored different machine learning algorithms and have performed an analysis. The classifier models used here in- clude Support vector machine, Random Forest and Logistic Regression out of which the maximum accuracy provides a prediction. This also integrated the models where users can enter their soil details and get a crop and fertilizer recommendation. The prediction made by machine learning algorithms will help the farmers to come to a decision which crop and fertilizer to grow to induce the most yield by considering factors like N, K, P, temperature, rainfall, location, etc.

Keywords Machine learning, Crop Prediction, Fertilizer Prediction, SVM, Logistic Regression, Random Forest.

1. INTRODUCTION

In India, one of the most significant pro- fessions is agriculture. It is the most diverse economic sector and is crucial to the overall progress of the nation. To meet the demands of 1.3 billion people, agriculture occupies more than 60% of the country's land. Conse- quently, implementing new agricultural tech- nologies is crucial[1]. This will be leading the farmers of our country towards profit. Prior crop and yield predictions were made based on the knowledge of farmers in a certain area. They don't know enough about the amount of soil nutrients like nitrogen, phosphorus, and potassium on the land, therefore they will prefer the older, more established, or more fashionable crop in the neighborhood just for their property[1]. Given the current circumstances, crop rotation is not being practiced, and the soil is not receiving enough nutrients, which reduces yield, causes soilpollution (soil acidity), and harms the toplayer[3].

It was created utilizing machine learning with all of these issues taken into consideration for the benefit of the farmer[2]. The agriculture industry has been trans- formed by machine learning (ML). In order to open up new potential for data-intensive science in the multidisciplinary field of agro- technology, machine learning, a subset of ar- tificial intelligence, has arisen along with big data technologies and high-performance computing[4]. Machine learning, for in- stance, is not a magic trick or a secret trick in the agricultural

industry; rather, it is a set of well-defined models that gather particular data and use particular methods to produce particular results[6][7].

It will recommend the most suitable crop and the fertilizer for particular land. Based on weather parameter and soil content such as Rainfall, Temperature, Humidity and pH according to the states in India[9]. It takes the required input from the farmers or sensors such as Temperature, Humidity, pH and area. This all-inputs data applies to machine learn-ing predictive algorithms like Random Forest and Logistic Regression[8] to identify the pattern among data and then process it as per input conditions. It recommends the crop for the farmer and also recommends the fertilizerto be add for the predicted crop[10].

2. LITERATURE SURVEY

S. Veenadhari [1] describes, in the world of agriculture, the climate is crucial. This year, the climate has been severely impacted by the rise in global warming, which has had a sig- nificant influence on agriculture. The farmers will be advised on what to harvest based on the crop production prediction using the pre- dictive analysis. Rakesh Kumar [2] describes, one of the key elements in the realm of agri- culture throughout the agro-based nations is economic growth and food security. Depend- ing on the environment, crop selection is a challenging challenge for agriculture plan- ning. It relies on a number of factors, includ- ing the weather, market prices, production rates, and governmental policies. Santosh Mahagaonkar [3] explain about Crop yield analysis, it is developing study area that in- cludes machine learning. A major problem in agriculture is yield prediction. Farmers have been forecasting crops for decades based on their knowledge and general assessment of aparticular crop.

PROPOSED WORK

Depending on the climatic conditions, includ- ing the soil's PH, temperature, humidity, and rainfall, the proposed method will suggest the best crop and fertilizer for a specific soil type. Python was used to create the system. The flow diagram in Fig. 1.

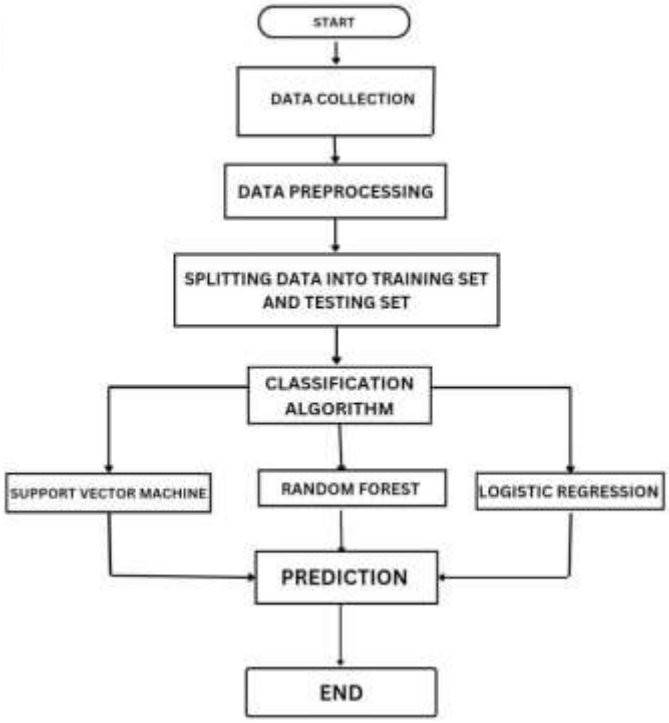
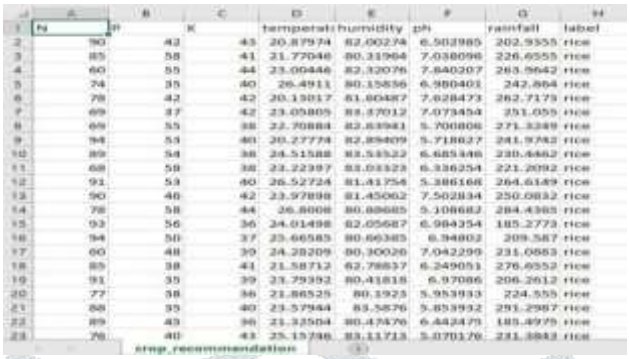


Fig . 1 Flow Diagram

3.1 DATA COLLECTION

Data collection is the most effective way to gather and analyses data from many sources, like Kaggle and others. in order to obtain a rough dataset for the system. These characteristics must be present in this dataset Temperature, Humidity, Soil pH, Rainfall, and Crop Data, NPK readings and fertilizer data are the characteristics that will be taken into account when predicting crops and ferti-lizer. Sample dataset in Fig. 2.



	N	P	K	temperature	humidity	ph	rainfall	label
2	90	42	43	20.87974	82.00274	6.502985	202.9355	rice
3	85	58	41	21.77046	80.31964	7.038096	226.6555	rice
4	60	55	44	23.00446	82.32076	7.840207	263.9642	rice
5	74	35	40	26.4911	80.15836	6.980401	242.864	rice
6	78	42	42	20.13017	81.60487	7.628473	262.7173	rice
7	69	37	42	23.05805	83.37012	7.073454	251.055	rice
8	69	55	38	22.70884	82.63941	5.700806	271.3249	rice
9	94	53	40	20.27774	82.89409	5.718627	241.9742	rice
10	89	54	38	24.51588	83.53522	6.685346	230.4462	rice
11	68	58	38	23.22397	83.03323	6.336254	221.2092	rice
12	91	53	40	26.52724	81.41754	5.386168	264.6149	rice
13	90	46	42	23.97898	81.45062	7.502834	250.0832	rice
14	78	58	44	26.8008	80.88685	5.108682	284.4365	rice
15	93	56	36	24.01498	82.05687	6.984354	185.2773	rice
16	94	50	37	23.66583	80.66385	6.94802	209.587	rice
17	60	48	39	24.28209	80.30026	7.042299	231.0863	rice
18	85	38	41	21.58712	82.78837	6.249051	276.6552	rice
19	91	35	39	23.79392	80.41818	6.97086	206.2612	rice
20	77	38	38	21.86525	80.1923	5.953933	224.555	rice
21	88	35	40	23.57944	83.5876	5.853932	291.2987	rice
22	89	43	38	21.32504	80.47476	6.442475	185.4979	rice
23	78	40	43	25.15746	83.11713	5.070176	231.3843	rice

crop_recommendation

Fig. 2 Sample Dataset

3.2 DATA PREPROCESSING

From the collected dataset, the model training must come before dataset preparation. Read- ing the obtained dataset is the first step in the data preprocessing process, which follows with data cleaning. When data is cleaned, some redundant attributes are removed from the datasets so that crop and fertilizer predic- tions can be made. Therefore, in order to im- prove accuracy, it is necessary to remove un- necessary attributes and fill up any missing values in datasets with undesirable nan val- ues, then indicate the target of the model. The techniques listed below, as displayed in theFig. 3

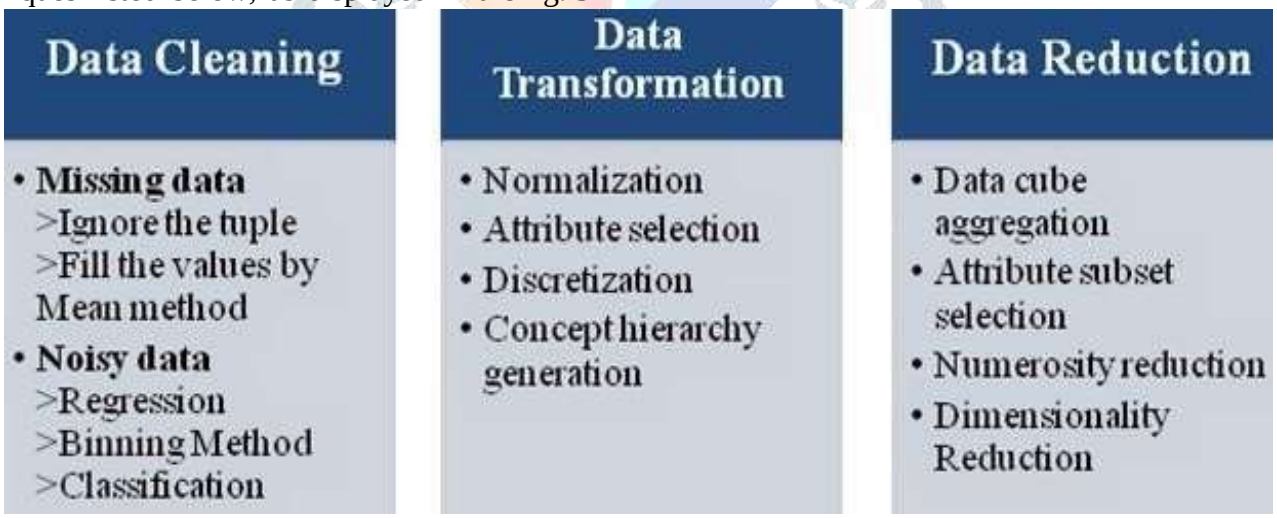


Fig. 3 Preprocessing methods

3.3 SPLITTING DATASET

While creating Machine Learning and Deep Learning Models may come across scenarios where want to do both training and well as evaluation on the same dataset. In such cases, may want to divide our dataset into different groups or sets and use each set for one task or spe- cific process (e.g., training). In such situations, have made use of training/test sets. The dataset divided into ratios of 70:30 and 80:20. Because the test and train sets have different requirements.

3.4 CLASSIFICATION

In this process we have implemented the machine learning algorithm such aa Support vector machine (SVM), Logistic Regression,Random Forest.

3.4.1 SUPPORT VECTOR MACHINE

One of the most well-liked supervised learning algorithms, Support Vector Ma- chine, or SVM, is used to solve Classificationand Regression problems.

Nevertheless, Machine Learning Clas- sification issues are where it is most fre- quently employed. In order to quickly catego- ries new data points in the future, the SVM algorithm aims to determine the optimum line or decision boundary that can divide n-di- mensional space into classes. A hyperplane is the name for this best-case decision bound- ary. SVM is used to choose the extreme vec- tors and points that contribute to the hyper- plane. The SVM approach is based on sup- port vectors, which are utilized to represent these extreme situations. Consider the dia- gram below Fig. 4, where a decision bound- ary or hyperplane is used to categories two distinct categories.

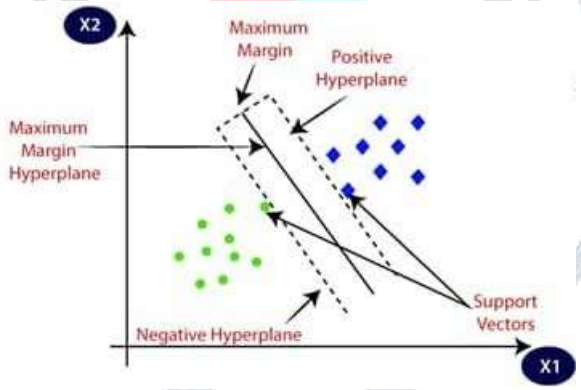


Fig. 4 Two distinct to categories hyperplane

3.4.2 LOGISTIC REGRESSION

Logistic regression is one of the Ma- chine Learning algorithms that is most fre- quently employed in the Supervised Learning category. Using a predetermined set of inde- pendent factors, it is used to forecast the cat- egorical dependent variable. A categorical dependent variable's output can be predicted using logistic regression.

The outcome must thus be a discrete or categorical value. Rather of providing the precise values of 0 and 1, it provides the prob- abilistic values that fall between 0 and 1. There are two possible outcomes: True or False, 0 or 1, or Yes or No.

Logistic regression and linear regres- sion are fairly similar, with the exception of how they are used. While logistic regression is used to address classification issues, re- gression issues are addressed by linear re- gression. In logistic regression, we fit a "S" shaped logistic function, which predicts two maximum values (0 or 1), rather than a re- gression line.

The logistic function's curve shows the possibility of several things, like whether or not the cells are malignant, whether or not a mouse is fat depending on its weight, etc. Us- ing both continuous and discrete datasets to classify fresh data, logistic regression is a crucial machine learning technique. When classifying observations using various sources of data, logistic regression may be used to quickly identify the factors that will work well. The logistic function is seen in thegraphic below Fig 5.

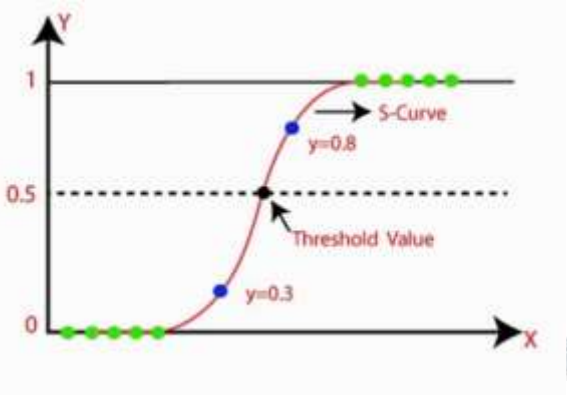


Fig. 5 Logistic Function

3.4.3 RANDOM FOREST

A supervised learning approach called random forest is employed for both classification and regression. But it is mostly employed for categorization issues. As is common knowledge, a forest is made up of trees, and a forest with more trees will be stur- dier. Similar to this, the random forest method builds decision trees on data samples, obtains predictions from each one, and then uses voting to determine the optimal option. Because it averages the results, the ensemble method which is superior to a single decision tree reduces over-fitting.

Random Forest's algorithmic operation, starting with the random sample selection from a given dataset is the first step. A deci- sion tree will then be built for each sample by this approach. From each decision tree, it will then get the projected outcome. Voting for each anticipated outcome will be done in this stage. Finalize the prediction result by choos-ing the one that received the most votes. Thebelow Fig. 6 illustrate its working

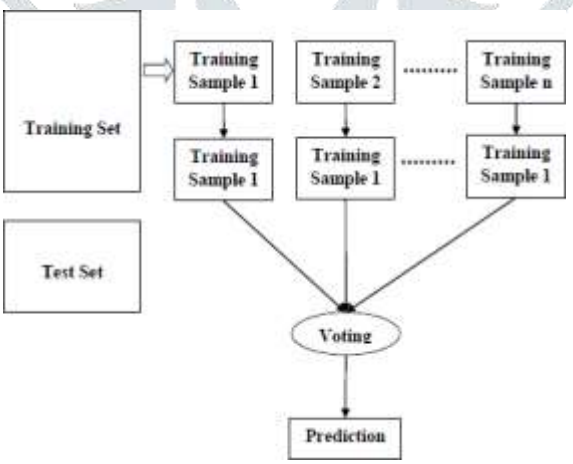


Fig. 6 Working of Random Forest

3.5 PREDICTION

PERFORMANCE METRICS

The Final Result will get generated based on the overall classification and pre- diction. The performance of this proposed approach is evaluated using some measureslike,

- Correlation matrix
- Accuracy and Error rate
- Then, we have to predict or to classify thesuitable crop and the fertilizer.

Correlation matrix

The Correlation matrix for crop andfertilizer dataset in below figure 7 and 8

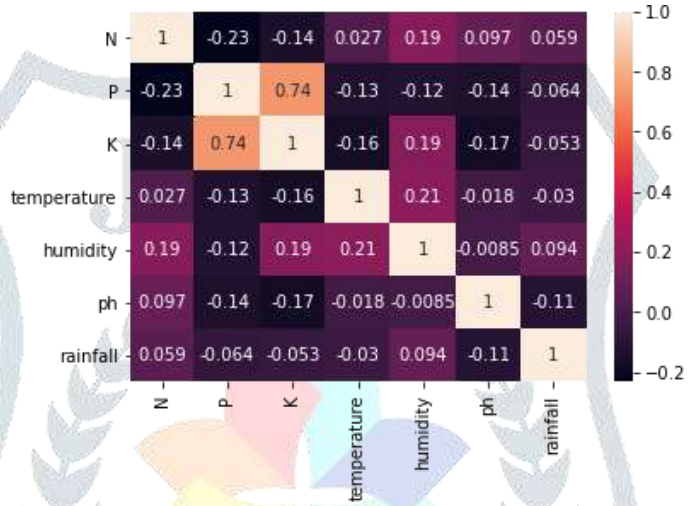


Fig. 7 Correlation matrix for Crop dataset

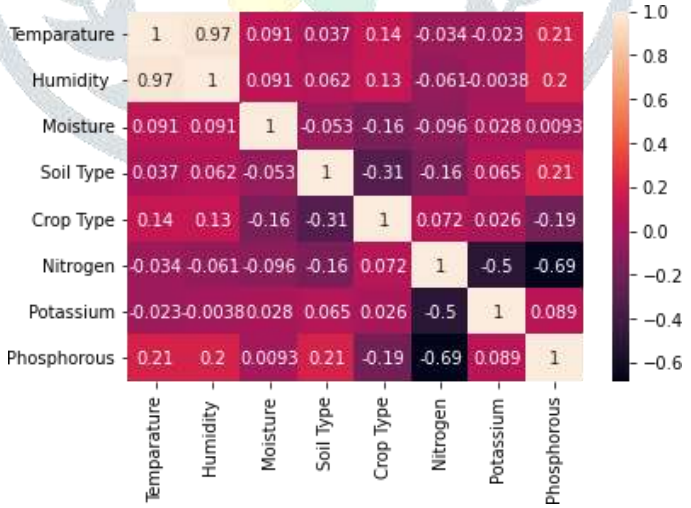


Fig. 8 Correlation matrix for Fertilizer Dataset

Accuracy and Error rate

Accuracy = Number of correct predictionsTotal number of predictions

Error rate = 1.0 – accuracy

Table 1: Accuracy and Error rate for ResultComparison to crop prediction

ALGORITHM	ACCURACY	ERROR RATE
SVM	0.97	0.03
LOGISTIC REGRESSION	0.95	0.05
RANDOM FOREST	0.99	0.01

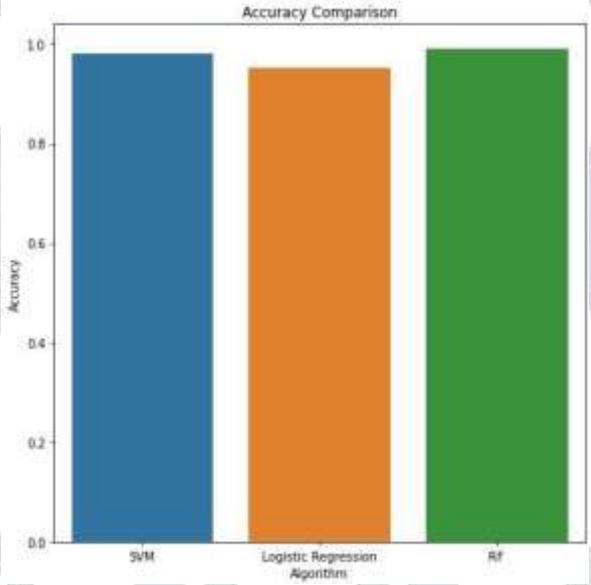


Fig. 9 Accuracy comparison for croppredict ion

Table 2: Accuracy and Error rate for ResultComparison to fertilizer prediction

ALGORITHM	ACCURACY	ERROR RATE
SVM	0.78	0.22
LOGISTIC REGRES-SION	0.98	0.02
RANDOM FOREST	0.96	0.04

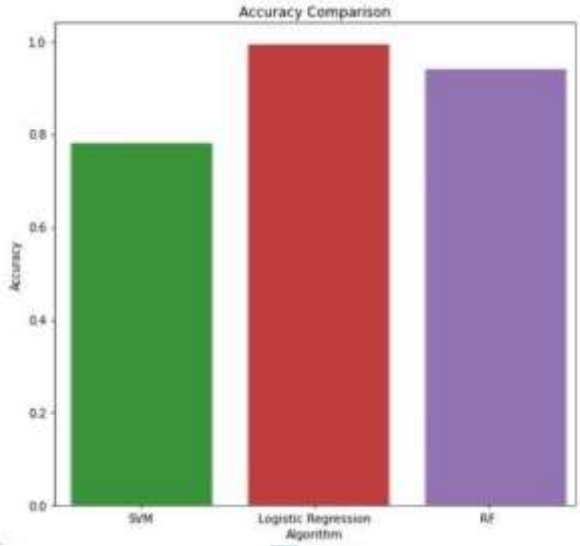


Fig. 10 Accuracy comparison for fertilizerprediction

3. CONCLUSION

Lack of profitability is one of the greatest problems that farmers today have in the agricultural sector. Crops are grown, but farmers don't obtain the right yield, which lowers their profit margin. In the agricultural department, yield prediction is important. Early yield forecasting is crucial, and should be based on variables like temperature, rain- fall, soil characteristics, etc. It produces a pre- cise crop output and provides the user with correct advice about the necessary fertilizer ratio based on the atmospheric and soil char- acteristics of the field, which improves the crop yield and boosts farmer income.

The future work is concentrated on providing the sequence of crops. To gather all necessary data, a land must provide its GPS location and access to the government's rain forecasting system, which can predict crops by simply providing GPS location. Addition- ally, a model for preventing food shortages and surpluses may be developed.

4. REFERENCE

[1] S. Veenadhari, B. Misra and C. Singh, "Machine learning approach for forecasting crop yield based on climatic parame- ters," 2014 International Conference on Computer Communication and Informatics, Coimbatore, India, 2014.

[2] A. Nigam, S. Garg, A. Agrawal and P. Agrawal, "Crop Yield Prediction Using Ma- chine Learning Algorithms," 2019 Fifth In- ternational Conference on Image Infor- mation Processing (ICIIP), Shimla, India, 2019.

[3] Bondre, Devdatta A and Santosh Ma- hagaonkar. "Prediction Of Crop Yield and Fertilizer Recommendation Using Machine Learning Algorithms." International Journal of Engineering Applied Sciences and Tech- nology (2019).

[4] Dhivya B, Manjula, Bharathi S and Madhumathi March 2017 "A Survey on Crop Yield Prediction based on Agricultural Data"International Conference in Modern Science and Engineering 2017

[5] Chandgude, A., Harpale, N., Jadhav, D., Pawar, P., & Patil, S. M. (2018). "A Review on Machine Learning Algorithms Used For Crop Monitoring Systems in Agriculture". In- ternational Research Journal of Engineering and Technology (IRJET) 2018

[6] Patil, A., Kokate, S., Patil, P., Panpatil, V., & Sapkal, R. (2020). "Crop Prediction using Machine Learning Algorithms". InternationalJournal of Advancements in Engineering & Technology 2020

[7] Omkar B,Nilesh M,Shubham G,Chandan M, and D.S. Zingade (2017), Crop Prediction System using Machine Learning Techniques,

International Journal of Advance Engineer-ing and Research Development

[8] A. Suresh, N. Manjunathan, P. Rajesh and
E. Thangadurai, "Crop Yield Prediction Us- ing Linear Support Vector Machine", Euro- pean Journal of Molecular & Clinical Medi- cine 2020.

[9] P. Tiwari and P. Shukla, "Crop yield pre- diction by modified convolutional neural net- work and geographical indexes", Interna- tional Journal of Computer Sciences and En-gineering 2018.

[10] R. Ghadge, J. Kulkarni, P. More, S. Nene and R. L. Priya, "Prediction of crop yield using machine learning", International Research Journal of Engineering and Tech- nology (IRJET)