INTERNATIONAL JOURNAL FOR INNOVATIVE RESEARCH IN MULTIDISCIPLINARY FIELD ISSN: 2455-0620 Volume - 7, Issue - 5, May – 2021
Monthly, Peer-Reviewed, Refereed, Indexed Journal with IC Value: 86.87 Impact Factor: 6.719
Received Date: 11/05/2021 Acceptance Date: 26/05/2021 Publication Date: 31/05/2021

Comparative Analysis For Crop Yield And Demand Predictive Using Machine Learning

[1] Sujal A. Shelke, [2] Nitin R. Chopde
[1]Research Scholar, Department of Computer Science and Engineering, G.H Raisoni University, Amravati , Maharashtra, India.
[2]Asst. Professor, Department of Computer Science and Engineering, G.H Raisoni University, Amravati, Maharashtra, India.
Email - [1]sujal23shelke@gmail.com, [2]nitin.chopde@raisoni.net

Abstract : *Agriculture is the most important field in the Indian economy, accounting for 18% of the country's GDP (GDP). Agriculture is the primary source of food in all countries, whether they are under-developed, developing, or developed. Farmers typically plan the cultivation process based on previous experience. They end up cultivating undesirable crops due to a lack of precise knowledge about cultivation. Predicting crop yields well in advance of harvest would assist farmers in taking appropriate marketing and storage measures. Machine learning is an important perspective for obtaining a real-world and operational solution to the crop yield issue. This paper focuses on the result analysis of a crop yield prediction model using different machine learning algorithms. The prediction made by a machine learning algorithm will help the farmers to decide which crop to grow to get the maximum yield by considering factors like soil moisture, temperature, rainfall etc. The K-nearest neighbor, SVM and Random forest algorithm will be used to predict the yield target/outcome and compare these algorithms to get the best accurate result.*

Key Words: *Crop yield prediction, Agriculture, Machine learning, K-nearest neighbor, Support vector machine.*

1. INTRODUCTION:

Agriculture is extremely important to the world economy. With the continued expansion of the human population, pressure on the agricultural system will increase. Agricultural yield in India is primarily determined by weather conditions. Agriculture is one among the substantial areas of interest to society since a large portion of food is produced by it. Currently, many countries still experience hunger because of the shortfall or absence of food with a growing population. Timely advice to predict future crop productivity and an analysis are required to assist farmers in maximising crop production.

Today Machine learning plays an important role in solving various classification, forecasting and analysis problems in various fields, like in analyzing soil type and further, it can be useful for crop prediction based on the previous crop sequences in the same farmland with the current soil nutrient information. SVM (Support vector machine) plays an important role in classification as it provides best partition among classes. Predicting a crop in advance necessitates a systematic study of vast amounts of data derived from various variables such as temperature, humidity, rainfall, crop name, soil quality, and so on. Because crop prediction involves a large number of databases, this prediction system is an ideal candidate for the application of a machine learning model. Many classification methods are also applied to get the maximum yield of crops.

2. AIM & OBJECTIVES :

- The objective of the project is to predict a crop yield well in advance. The crop yield prediction incorporates forecasting the yield of the crop from past historical data by classification algorithms like support vector machine, Random forests, K-nearest neighbor.
- Demand for crops can be predicted based on past data and current climate and soil data with the help of Random forests algorithm on past rain data and atmospheric parameters like temperature, humidity etc.
- Demand predictions can lead to supply predictions, which are interdependent; these demands can be based on current rain data and atmospheric conditions, which leads to crop production and generates revenue.

3. RELATED WORK:

Machine learning is a branch of computer science that is used to create algorithms that have self-learning properties, i.e. learning that is done by the machine itself, hence the term "Machine Learning." It is viewed as one of the

significant areas of Artificial Intelligence. To show intelligence, the machine must interpret and analyse the data. Apart from just following the instructions given on the data, after analyzing the result data. This is the thing that machine learning algorithms do. Machine learning centers on the development of computer programs that can get data and utilize it to learn for themselves. In order to search for patterns in data, the path to learning begins with perceptions of information, such as direct experience or instruction. It helps to make better decisions in the future based on the examples that we give. The key is to let the computer learn on its own, without human intervention or assistance, and then regulate actions accordingly. It is a major field of computer science that is used in a variety of advanced technological development programmes all over the world. In another research, the author K. Kaur utilizes machine learning in different applications in Indian agriculture [5].

In this paper the various applications of machine learning techniques in agriculture have been listed, such as Crop Selection and Crop Yield Prediction, Weather Forecasting, Smart Irrigation System, Crop Disease Prediction, Deciding the Minimum Support Price. These techniques will enhance the productivity of fields along with a reduction in the input efforts of the farmers. Along with the advances in machines and technologies used in farming, useful and accurate information about different matters also plays a significant role in it [4].

The authors S. Ying-xue, X. Huan, and Y. Li-jiao in their research work describe a crop model as a computer program that mathematically describes and models the principles of harvest growth and can be utilized to quantitatively and dynamically clarify the procedure of product development, improvement, yield, and response to ecological change [6].

Crop models can be classified as harvest factual models and crop simulation models (or crop growth models) in light of the fundamental numerical technique for the displaying. Furthermore, in another paper, the authors used Auto-regressive Integrate Moving Average (ARIMA) and The Support Vector Machines (SVM) [7].

4. PROPOSED METHODOLOGY:

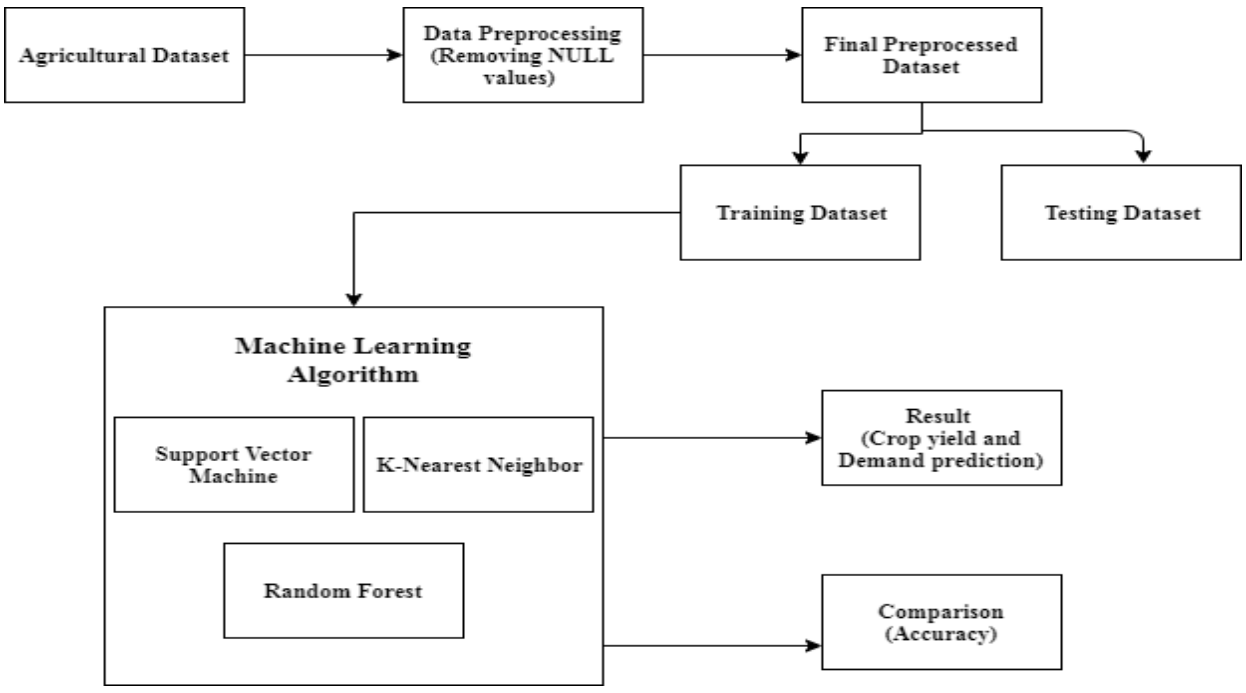


Fig- 1.1: Block Diagram Of Crop Yield And Demand Prediction

The above architecture clearly explains how the system's components communicate with one another, beginning with data preprocessing. The crop yield can be estimated using this proposed framework. This model gives a clear picture of a large amount of data capture and data preprocessing to remove unwanted data such as NULL, etc. presented in it. During the preprocessing step, we split the dataset into training and testing datasets.

Train the dataset using appropriate supervised learning algorithms to detect the crop yield present in the dataset. Apply machine learning techniques to find crop yield for any new data that has appeared in the data set. After this data acquisition, suitable machine learning algorithm must be applied to compute efficiency and capability of the model.

Here we have applied various machine learning algorithms like support vector machine, K-nearest neighbor, random forest etc. Metrics such as accuracy and precision will be calculated for the proposed model.

4.1 Support Vector Machine

"Support Vector Machine" (SVM) is a supervised machine learning algorithm which is a very useful technique for data classification. However, this learning algorithm can also be used for regression challenges. A classification task usually involves dividing data into training and testing sets.Each instance in the training set has one "target value" (i.e. the class labels) and several "attributes" (i.e. the features or observed variables). The goal of SVM is to build a model (based on training data) that predicts the target values of test data given only the test data attributes.

4.2 K-Nearest Neighbor

K-Nearest Neighbors (KNN) algorithm is one of the simplest, easy to understand, versatile and one of the topmost machine learning algorithms. KNN is a non-parametric supervised learning algorithm. It is also an instance-based learning or a lazy algorithm. The algorithm uses the training instances to spit out an answer when a query to the database is made.That is why, for KNN the training phase is very fast compared to other classifier algorithms. The testing phase, on the other hand, becomes slower and more expensive in terms of time and memory. The testing phase, on the other hand, becomes slower and more expensive in terms of time and memory.

4.3 Random Forest

Random forests are a set of tree predictors in which each tree is determined by the values of a random vector sampled independently and with the same distribution for all trees in the forest. Random Forest is a highly flexible and simple to use machine learning algorithm that produces excellent results most of the time even without hyper-parameter tuning. It is also one of the most widely used algorithms due to its simplicity and ability to perform both classification and regression tasks.

4.4 Agricultural Datasets

The dataset consists of factors like temperature, rainfall, moisture, humidity, alkaline, sandy . The datasets have been obtained from the Kaggle website and other different websites. The data set has 5000 instances or data that have been taken from the past historic data. It includes 8 parameters or features like the temperature, rainfall, moisture, humidity, alkaline, sandy etc.

5. ANALYSIS & RESULT:

These are complete datasets we have obtained from Kaggle and other different websites from all over India. This dataset has 8 parameters like moisture, rainfall, average humidity, mean temperature, max temperature, min temperature, alkaline, sandy and the predicted crop yield value. It contains nearly 5000 records. Table 1 shows the description of the agriculture dataset. Table 2 shows the crop yield predicted value and the crop yield and demand predicted output.

0,1,2,3,4 are the 5 values which are predicted by applying the different algorithms in the dataset. Here we are applying the 3 algorithms which support vector machine, K-nearest neighbor and random forest. The Table-2 shows the output in the system will predict the crop yield and the demand in high, moderate and low format.

Table-5.1: Complete Dataset

Sr	Moisture	Rainfall	Average_Humidity	Mean_Temp	Max_Temp	Min_Temp	Alkaline	Sandy	Crop_Yield
1	12.80168453	0.0123605	57	62	71	52	0	1	2
2	12.85165378	0.00417157	57	58	73	43	0	1	0
3	12.7767735	0	56	58	69	46	0	0	4
4	12.94200101	0.03174683	62	57	70	43	0	1	0
5	12.98465248	0	65	56	70	42	0	0	1
6	12.96447065	0.02719149	65	58	70	46	1	0	0
7	12.73799817	0.02682104	61	56	70	42	0	0	1
8	12.81938179	0.01028368	58	57	72	42	0	0	1
9	12.88390946	0.02046472	63	61	76	45	0	0	0
10	12.78451285	0.06005408	62	59	71	47	0	1	4

Table-5.2 : Crop Yield And Demand Predicted Output With Values

Crop Yield Predicted Value	**Crop yield and Demand predicted output**
0	Crop yield will be very low and Demand will be extremely high
1	Crop yield will be moderately low and Demand will be high
2	Crop yield will be moderate and Demand will also be moderate
3	Crop yield will be high and Demand will be low
4	Crop yield will be extremely high and Demand will be very low

Table -5.3: Classification Report

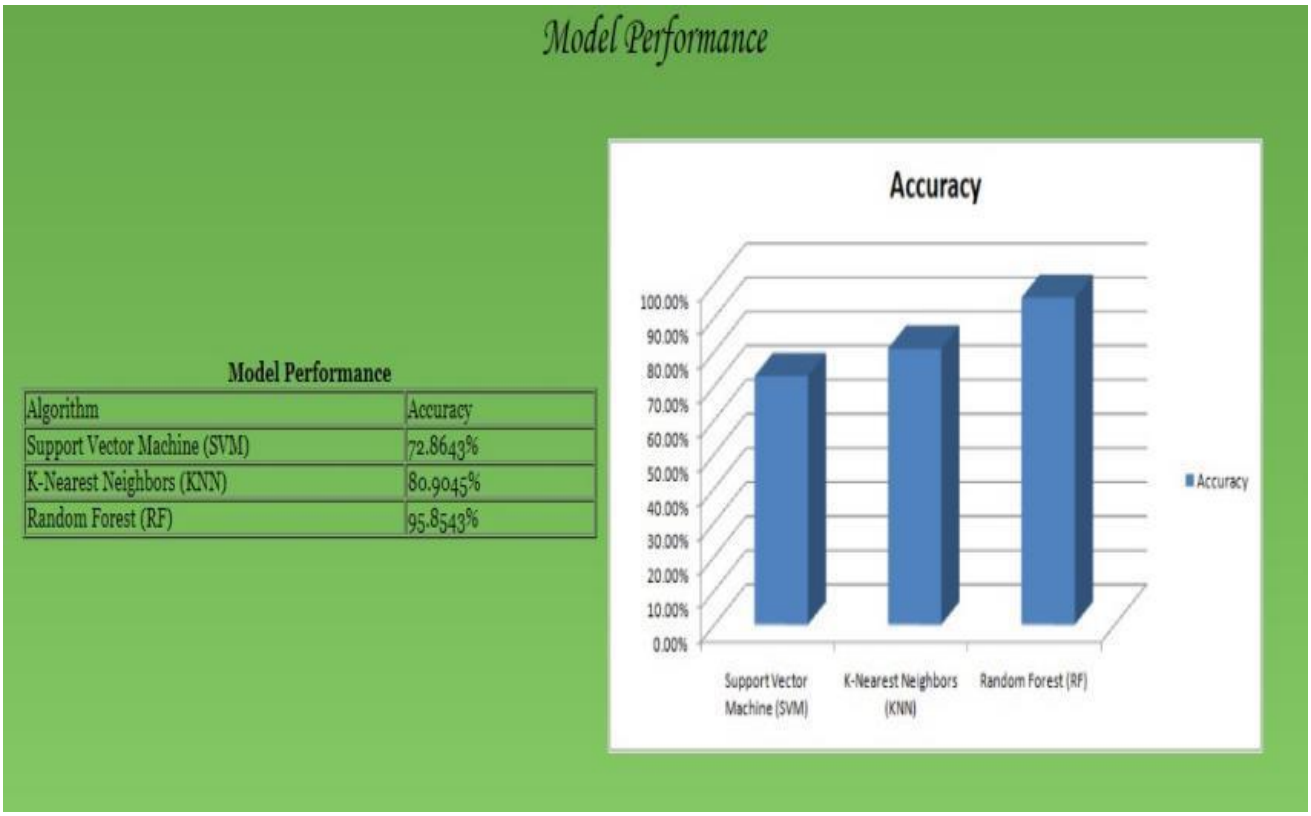
Classification Report

Algorithm	Report
Support Vector Machine (SVM)	precision recall f1-score support 0 0.73 0.68 0.70 253 1 0.83 1.00 0.90 142 2 0.64 0.75 0.69 126 3 0.00 0.00 0.00 13 4 0.71 0.66 0.69 262 accuracy 0.73 796 macro avg 0.58 0.62 0.60 796 weighted avg 0.72 0.73 0.72 796
K-Nearest Neighbors (KNN)	precision recall f1-score support 0 0.84 0.74 0.79 253 1 0.81 0.95 0.87 142 2 0.85 0.79 0.82 126 3 0.82 0.69 0.75 13 4 0.77 0.81 0.79 262 accuracy 0.81 796 macro avg 0.82 0.80 0.80 796 weighted avg 0.81 0.81 0.81 796
Random Forest (RF)	precision recall f1-score support 0 1.00 0.91 0.95 253 1 0.87 0.96 0.91 142 2 0.98 1.00 0.99 126 3 0.92 0.92 0.92 13 4 0.97 0.98 0.98 262 accuracy 0.96 796 macro avg 0.95 0.96 0.95 796 weighted avg 0.96 0.96 0.96 796

Table -5.4 : Confusion Matrix

Confusion Matrix

Algorithm	Matrix
Support Vector Machine (SVM)	[[171 18 17 0 47] [0 142 0 0 0] [5 5 94 0 22] [0 0 13 0 0] [59 7 23 0 173]]
K-Nearest Neighbors (KNN)	[[187 16 5 0 45] [1 135 1 1 4] [7 4 100 0 15] [1 0 2 9 1] [26 12 10 1 213]]
Random Forest (RF)	[[230 17 3 0 3] [0 137 0 1 4] [0 0 126 0 0] [1 0 0 12 0] [0 4 0 0 258]]



Model Performance

Algorithm	Accuracy
Support Vector Machine (SVM)	72.8643%
K-Nearest Neighbors (KNN)	80.9045%
Random Forest (RF)	95.8543%

Chart-5.1: Accuracy Comparison Among Algorithms For Crop Yield And Demand Prediction
The above chart shows the comparison of accuracy between the support vector machine ,K-nearest neighbor and the random forest algorithms.

6. CONCLUSION:

Agriculture is a fundamental aspect of modern civilization. With increasing world hunger and economic breakdown, the proper selection of crops emerges as a massive factor in this. But this is lacking in using new technologies of machine learning. As a result, our farmers should be familiar with all of the latest machine learning and other techniques. These techniques help in getting the maximum yield of crops. Many techniques of machine learning are applied to agriculture to improve the yield rate of crops. We can also get the accuracy of yield by checking for different methods. This paper helps to predict the crop yield and demand and compare the accuracy of the different machine learning algorithms.

7. FUTURE SCOPE

We hope that in the future, this model will be implemented with a much more efficient dataset for a specific piece of land that contains information such as different soil properties, soil pH, different mineral percentages, and so on.. So that no agricultural plot is wasted by harvesting a less productive crop. We want this model to be used all over the world to help the agricultural sector grow. Both researchers and entrepreneurs may be interested in this field. In the future, we hope to create a platform for all of the farmers who will be using this model to share the predictions on their land with other farmers all over the country. This will let a farmer in one country to know the prospect of farming in another part of the world; down to a specific geographical unit.

REFERENCES:

1. Monali Paul, Santosh K. Vishwakarma, Ashok Verma, “Analysis of Soil Behavior and Prediction of Crop Yield using Data Mining approach”, 2015.
2. N. Heemageetha, “A survey on Application of Data Mining Techniques to Analyze the soil for agricultural purpose”, 2016IEEE.
3. Y. Jeevan Nagendra Kumar, Dr. T . V. Rajini Kanth, “GISMAP Based Spatial Analysis of Rainfall Data of Andhra Pradesh and Telangana States Using R”, International Journal of Electrical and Computer Engineering (IJECE), Vol 7, No 1, February2017, Scopus Indexed Journal, ISSN: 2088-8708.

4. Jain, N., Kumar, A., Garud, S., Pradhan, V., and Kulkarni, P. (2017). Crop selection method based on various environmental factors using machine learning.
5. Kaur, K. (2016). Machine learning: applications in Indian agriculture Int. J. Adv.Res. Comput. Commun. Eng.(IJARCCE), 5(4).
6. Su, Y.-x., Xu, H., and Yan, L.-j. (2017). Support vector machine-based open crop model (sbocm): Case of rice production in china. *Saudi journal of biological sciences*, 24(3):537–547.
7. Sujjaviriyasup, T. and Pitiruek, K. (2013). Agricultural product fore-casting using machine learning approach. Int. Journal of Math. Analysis, 7(38):1869–1875.
8. Narayanan Balkrishnan and Dr. Govindarajan Muthukumarasamy. "Crop production Ensemble Machine Learning model for prediction." International Journal of Computer Science and Software Engineering (IJCSSE).
9. Nilesh Dumbre, Omkar Chikane, Gitesh More, "System for agriculture recommendation using data mining." International Education & Research Journal [IERJ] E-ISSN: 2454- 9916 | Volume: 1 | Issue: 5 | Dec 2015.
10. Anshal Savla, Himtanaya Bhadada, Parul Dhawan, Vatsa Joshi, "Application of Machine Learning Techniques for Yield Prediction on Delineated Zones in Precision Agriculture,". May 2015.
11. Zelu Zia "An Expert System Based on Spatial Data Mining used Decision Tree for Agriculture Land Grading." Second International Conference on Intelligent Computation Technology and Automation. Oct10-11, China
12. Ankit Narendrakumar Soni "Crack Detection in buildings using convolutional neural Network" Volume-2, Issue-6 (June-2019), Journal For Innovative Development in Pharmaceutical and Technical Science ISSN (O):- 2581-6934.
13. Mrs.K.R.Sri Preethaa, S.Nishanthini,D.SanthiyaK.VaniShree, "Crop Yield Prediction," International Journal On Engineering Technology and Sciences – IJETS™ISSN(P): 2349-3968, ISSN (O):2349-3976 Volume III, Issue III, March- 2016.
14. Aditya Shastry, H.A Sanjay And E.Bhanushree, "Prediction of crop yield using Regression Technique," International Journal of computing r12 (2):96-102 2017, ISSN:1816-914] E.14]E. Manjula, S. Djodiltachoumy, "A Model for Prediction of Crop Yield," International Journal of Computational Intelligence and Informatics, Vol. 6: No. 4, March, 2017.
15. Su, Y.-x., Xu, H., and Yan, L.-j. (2017). Support vector machine-based open crop model (sbocm): Case of rice production in china. Saudi journal of biological sciences, 24(3):537–547.
16. Sujjaviriyasup, T. and Pitiruek, K. (2013). Agricultural product fore-casting using machine learning approach. Int. Journal of Math. Analysis, 7(38):1869–1875.