Crops Prediction Based on Environmental Factors Using Machine Learning Algorithm

Journal of Development Economics and Management Research Studies (JDMS)
A Peer-Reviewed Open Access International Journal
ISSN: 2582 5119 (Online)
Crossref Prefix No: 10.53422
09(11), 127-137, January-March, 2022

https://www.cdes.org.in/about-journal/

Crops Prediction Based on Environmental Factors Using Machine Learning Algorithm

Dr. Poorna Shankar[1], Dr. Prashant Pareek[2], Ms. Urvi Patel[3], and Mr. Canny Sen[4]

Abstract

India is an agricultural country, much of the economy is dependent on productivity growth. Agriculture is heavily dependent on rainwater and depends on various soil conditions, namely nitrogen, phosphorus, potassium, and climates such as temperatures and rainfall. The growth of agricultural technology will increase crop production. Machine learning is a promising area for research to anticipate yield based on data patterns. The proposed learning algorithms apply to the machine learning algorithms: Random Forest, Logistic Regression, Decision Tree, and Support Vector Machine. Predictions of plants that are most relevant to the current environment are being made. This work gives producers a strong prediction of planting what types of crops in their area on the farm according to the above-mentioned parameters to grow a smart agricultural product. four different algorithms are applied in this project system. With the help of the ROC-AUC-SCORE, the accuracy of all the models is compared and other factors like precision, recall, F1 score, and support are also compared. And from all these results we can know which model is perfect and from that, we can know which crop is suitable for the given soil and climatic condition.

Keywords: Crop Prediction, Logistics regression, K-means clustering, exploratory analysis, Random Forest, Decision tree, Support vector machine.

1. Introduction

Ability to construct a single assessment model to indicate the most favourable plants to grow based on current climate and soil conditions. Direct farming is the adoption of more specialized methods that use technology to meet the needs of individual sites and plants. Direct

[1] Professor, Keystone Global, Ahmedabad, Gujarat.
[2] Assistant Professor, Shanti Business School, Ahmedabad, Gujarat.
[3] Pursuing MSC in Big Data, FOM University, Germany.
[4] Pursuing MSC in Big Data, FOM University, Germany

agriculture helps farmers to live a debt-free life as there is less agriculture and losses and the overall environmental impact is also lower. Direct farming will be used for this project. With this technique, the farmer can estimate the yield. Thereby, the farmer can get a livelihood. India is an agricultural country. So, if every farmer is healthy, the Indian economy will also be strong. By using this method, the farmer does not know which crop is more profitable from the soil and nature point of view. So, the farmer does not give up and the farmer earns good money so that the farmer does not commit suicide and with this technology, we can reduce farmer suicide cases. So, this project should be done keeping in view the situation of the farmer. The project predicts high yields of low-cost crops. Provides a system that allows farmers to access relevant information. This project will help to create an assessment model that will provide the most effective solution. Determines additional variations in yield defined by soil and climatic variations.

The agricultural sector is undergoing a transformation driven by new technologies, which looks very promising, as it will enable us to take this major sector to the next level of agricultural production and agriculture. Precision agriculture, when the necessary inputs are needed and when needed, is the third wave of the modern agrarian revolution (the first is mechanization and the second is the green revolution and genetic modification), and systems are being improved by increasing agricultural knowledge these days. Availability of large amounts of data. This research work is the result of a crop planning program to increase productivity based on key technologies: the Internet of Things and Machine Learning Techniques. Machine learning techniques make plant estimates based on data. The use of this technology helps the farmer to get a better agricultural product. The main goal of agriculture is to improve crop yields with low maintenance costs and low environmental pollution. Potential growth and yield depend on many different production characteristics such as climate, soil characteristics and irrigation and fertilizer management.

Agriculture is the backbone of any economy. In a country like India where food demand is increasing due to population growth, there is a need to make progress in agriculture in line with demand. Agriculture is considered to be the main and most important tradition practised in India since ancient times. In this project, we will use the extensive database available on the Coley website to create a model that can predict which crops will adapt to the given soil and climate change. The database contains a variety of plant extracts, including temperature, humidity, pH, precipitation, nitrogen (N), potassium (P), and phosphorus (K). After collecting the data, we clean the data to remove the lost values and if there are any vendors find them and manage them and we can do some data analysis to understand the data more deeply. We extract the feature and simulate the estimate using feature engineering. The goal of this project is precision farming. We need to improve productivity by understanding the climate and soil requirements of plants. It helps us deal with unpredictable weather. Compilation and development of crop yield estimates using data analytics to help increase yields and subsequent profitability in agricultural production.

1.1 Machine Learning

Machine learning is the most widely used method in agricultural matters. It is used in the analysis of big data collections and information categories. Creating useful fields and patterns. The general purpose of the machine learning method is to obtain data from the data collection then converts it into an understandable framework for continuous use. The main purpose of this paper is to develop a system that can predict the type of plant depending on soil and climate factors. As the population grows in today's world, as the years go by, this is expected to be in the billions and we need increased product production to cater to those billions of people. the population is growing, too, on the other hand, the agricultural sector is declining

due to several factors. such as major industrial development, commercial market development in residential buildings built on the farm; Therefore, supply with a growing number of consumers, there is a need to improve actual production can be obtained by using the most appropriate and most important object needed everyday life is clever farming.

2. Research Methodology

The research approach focuses on the latest technology that can be used to provide an alternative to the current approach to agriculture. The study is divided into 3 sections. Section 1 deals with machine learning and precision agriculture, section 2 deals with outcomes and consequences, section 3 deals with the future and scope of the future.

2.1 Machine learning cognitive technology

According to Mission Learning is a branch of artificial intelligence and computer science responsible for the development of algorithms that represent independent learning materials. With the help of accurate and efficient machine learning systems capable of exploring a very broad set of tasks are developed to solve day-to-day tasks. Scientists can use computer simulations to perform early plant experiments to determine how certain species can adapt to different climates, soil types, climatic patterns, and more. These digital experiments do not replace field testing, but allow plant growers to more accurately predict crop performance.

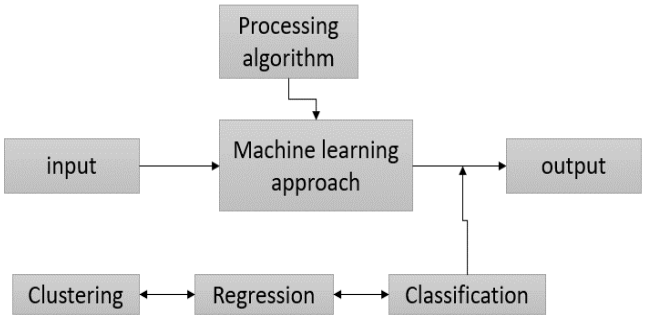

Figure 1 Machine learning data processing diagram

Information about the plant species to be tested is used as input and transmission through a surveyed, supervised or non-monitored machine learning system such as Random Forest, Logistics regression, Decision tree and vector support machine etc. The method analyses the input to extract relevant features and information related to the issue. Depending on the variety and functions set, the processing algorithm performs the data analysis and provides the output of possible split or retrospective output.

2.2 Impact of Precision Agriculture

Using agricultural understanding technology can help determine the best crop selection for different climatic conditions and best suits the needs of farmers. This can be achieved by analysing and comparing information about seed varieties, climate, soil types, specific attack's location, probability of disease and data about what worked best, annual results, current market

trends, prices and consumer needs. Farmers can then decide how to increase their crop yields. The speed at which machine learning technology develops will show that the agricultural industry is on the verge of technological change under artificial intelligence as its driving force.

2.3 Machine Learning Algorithm for Prediction

Machine learning is often used in agricultural problems. It is used in the analysis of big data collections and information categories to create useful fields and patterns. The general purpose of machine learning is to extract data from data collection and convert it into an understandable framework for continuous use. Based on accessible information, this document analyses the crop yield form. To increase crop production, a machine learning method was used to predict crop yields. This reflects the distribution of the predicted yield forecast. As shown in the above scenario, devices are used on the farm to obtain information connected to humidity, temperature, rainfall and ph. Logistics Regression algorithms are used to identify data obtained. The expected result indicates any soil that may be suitable for certain crops and the condition of groundwater.

A. Overview of Data

We get information from various places and organize data sets in this category. And for analytics, these data sets are used. Online tools such as Data.gov.in and Kaggle.com are also used to create accurate information. Time shows that the height of the plant indicates the temperature required for the plant in months, the minimum and the size of the feature. N, P, K values are plant-specific fertilizers, low and high PH values, maximum and minimum rainfall requires for crops and soil moisture levels. Use the data to estimate only for limited crops like fruits, grains, beans etc. and the farmer must be financially strong to use this project or this application. Also, for the weather, we should use only two factors namely temperature and rainfall. For soil, we should use only certain elements like nitrogen (N), potassium (P), phosphorus (K), moisture, pH etc. The dataset has 2,200 records and 8 attributes.

B. Logistics Regression

The logistic regression algorithm in ML is taken from the general logistic regression model in statistics. In modelling yields from a cropping system based on agronomy (upper and lower), the logistics function is estimated as follows:

In this case, we are modelling the potential for an input (X) (yield from four different crop systems) class (Y = 1 (Highlands)), and we can formally write it as:

$P(X) = P(Y = 1|X)$.

The logistic regression algorithm (LR) was used as the classification method (linear method), but the estimates were changed using the logistic function. The function can be called:

$$P(X) = e\frac{\beta 0 + \beta 1 * X}{1 + e^{\beta}0 + \beta 1 * X}$$

C. Random Forest

Random Forest is a supervised learning algorithm used for both classification and regression. The Random Forest Algorithm builds a decisive tree on different data models and then estimates the data from each subset and provides a better solution for the system through polling. Used the bagging method to train random forest data. The bagging method is to study different models and improve the result of the system. We used a random forest algorithm to

obtain high accuracy, which gives the accuracy estimated by the model and the actual results of the estimate in the dataset. The decisive tree is formed from data and tree patterns in the random forest provides an estimate from each family and select the best solution by voting to provide better accuracy to the model. This will give the correct results for the system.

D. Decision Tree

The Decision Tree is a data structure constructed using nodes (values or conditions) and margins (connecting all nodes). The decision tree is constructed based on a dataset that contains the characteristics or features that characterize the raw data for each record. Each node in a tree can be a decision-making node or a leaf node that delivers a result. The ID3 algorithm before the C4.5 algorithm tries to build the decision tree as small as possible. It uses the entropy of each feature to determine which edge to follow. Since ID3 is a greedy top-down approach, it identifies the feature with extreme values (highest data gain features). It helps to identify which features are from the most homogeneous branches.

E. Support Vector Machine

Like a popular machine learning algorithm, SVM is a new generation learning system. Based on recent advances in statistical learning theory. This VC Absorbs the principle of the magnitude and principle of the minimum structural risk to the objective function and then find the partition hyperplane that can meet the square requirement. The important advantage of SVM is that it can be theoretically analysed using cutting concepts from computational learning theory and achieve cutting-edge performance. Recently, it has also been applied to several real-world issues such as handwriting recognition, data retrieval, and biomedical data classification. In this paper, SVM is introduced to classify farm data to improve farm data classification performance.

3. Implementation

After collecting the features from the submitted data, we can easily estimate Depending on the type of crop on the characteristics we consider. Almost After evaluating the effectiveness and efficiency of our experiment, we can evaluate the appropriate crops for the given attributes.

3.1 Exploratory Data Analysis

We did exploratory data analysis to find the correlation of data and to understand how was the data distributed.

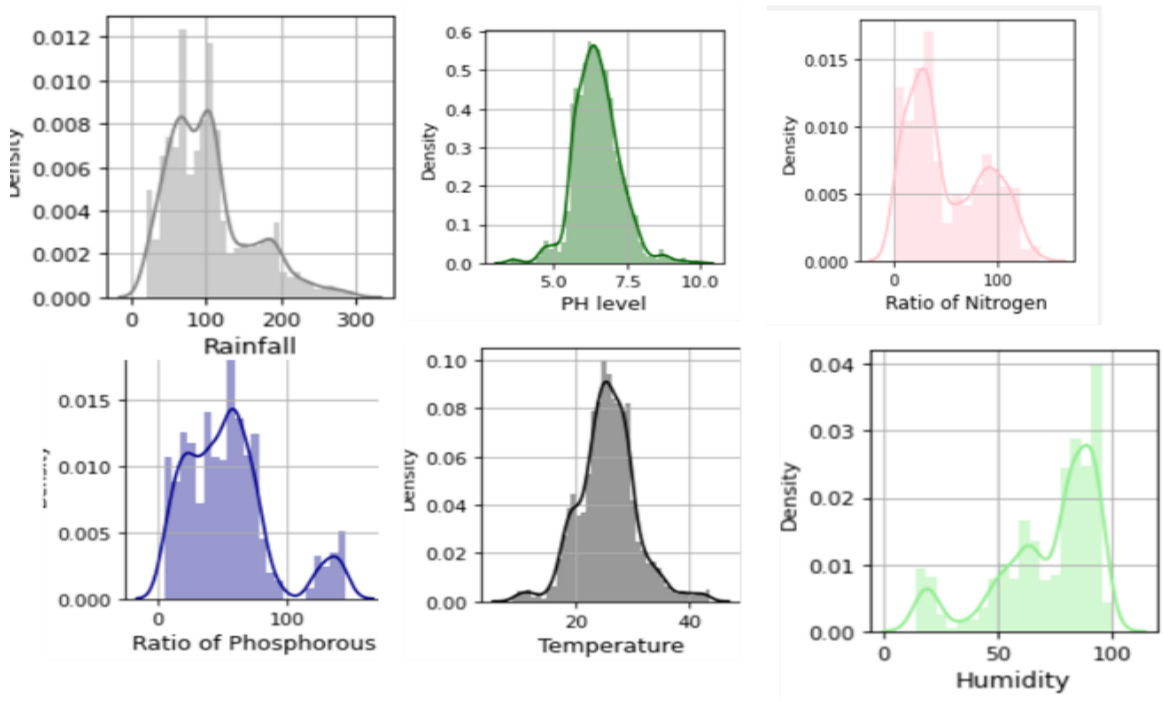

Figure 2 Distribution for Agricultural Conditions

These are distribution charts for all climate and soil conditions. From this chart, we know that some crops require high amounts of phosphorus and potassium. Because the value of phosphorus and potassium charts in the charts that have been found is very high. This chart shows that some crops require very high and very low temperatures. And some crops require less and more pH in the soil.

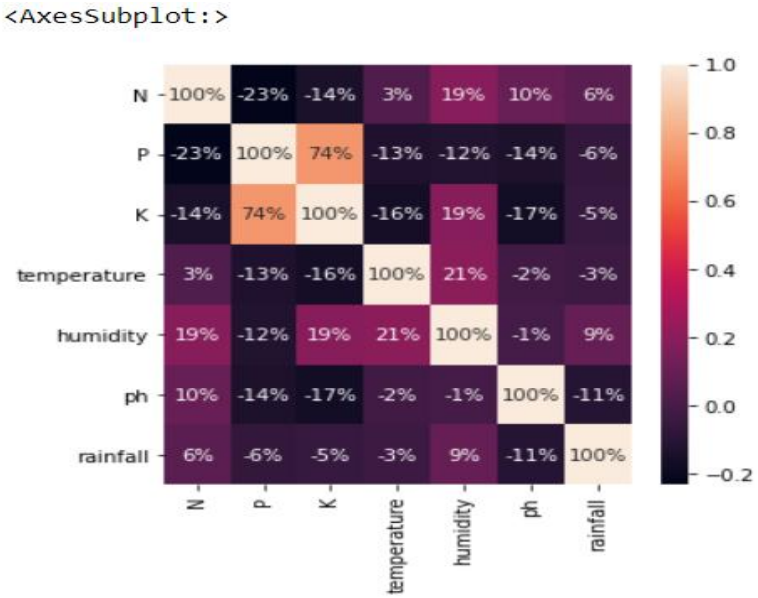

Figure 3 Correlation of attributes

We found the correlation between the data using a different combination of attributes and graphs. Figure. 3 shows how each attribute is correlated to attributes.

3.2 Classification Algorithm

After understanding data properly, we train our prediction algorithm one by one and obtain precision, recall, F-score and accuracy. The random forest classifier showed excellent results with an accuracy of 100%, whereas, the accuracy of Logistics regression was 97%, the accuracy of the Decision tree was 99% and Linear Support Vector Classifier was 98%. Table 1 shows the error rate and accuracy of each algorithm.

	Error rate	Accuracy
Random Forest	0.00	1.00
Logistics Regression	0.03	0.97
Decision Tree	0.01	0.99
Support Vector Machine	0.02	0.98

Table 1 Results of algorithms

Table 1 shows the accuracy Score of various classification Algorithms are compared and the best algorithm based on the accuracy score is chosen for the system.

4. Results and Model Evaluation

With the help of this confusion matrix, we can know whether our model is accurate or not. With the help of a classification report, we know the value of precision and recall. We can call our model good only if the value of precision and recall of our model is good. In this project, the value of recall and precision for 22 classes of our dataset is very high which means that our model is very accurate.

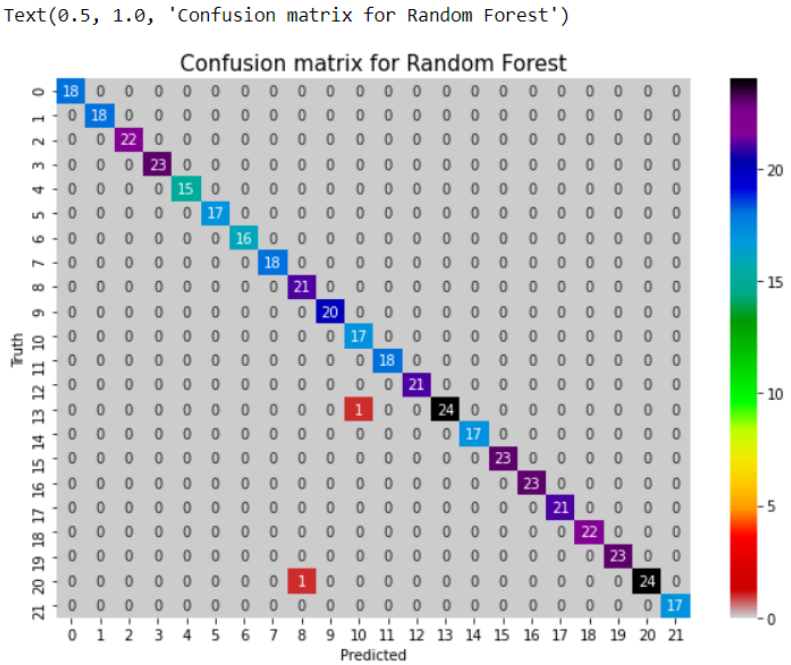

Figure 4 Confusion matrix for Random Forest

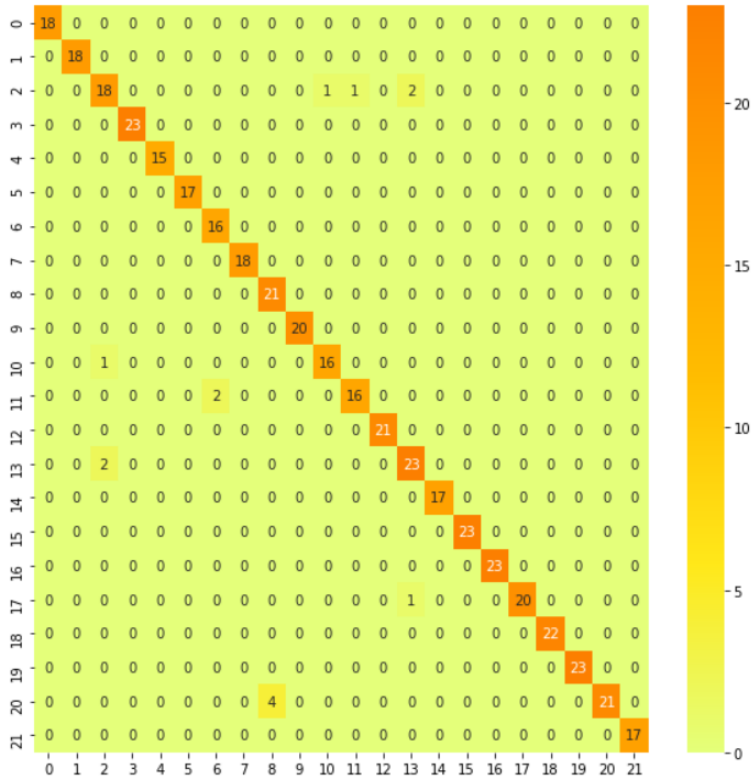

Figure 5 Confusion matrix for Logistics Regression

```
Text(0.5, 1.0, 'Confusion matrix for Support Vector Machine')
```

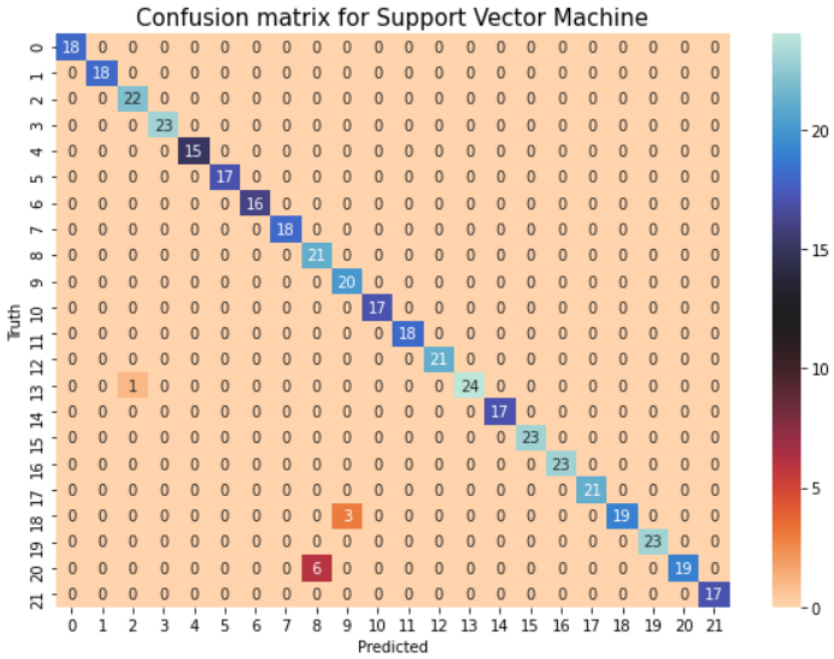

Figure 6 Confusion matrix for Support Vector Machine

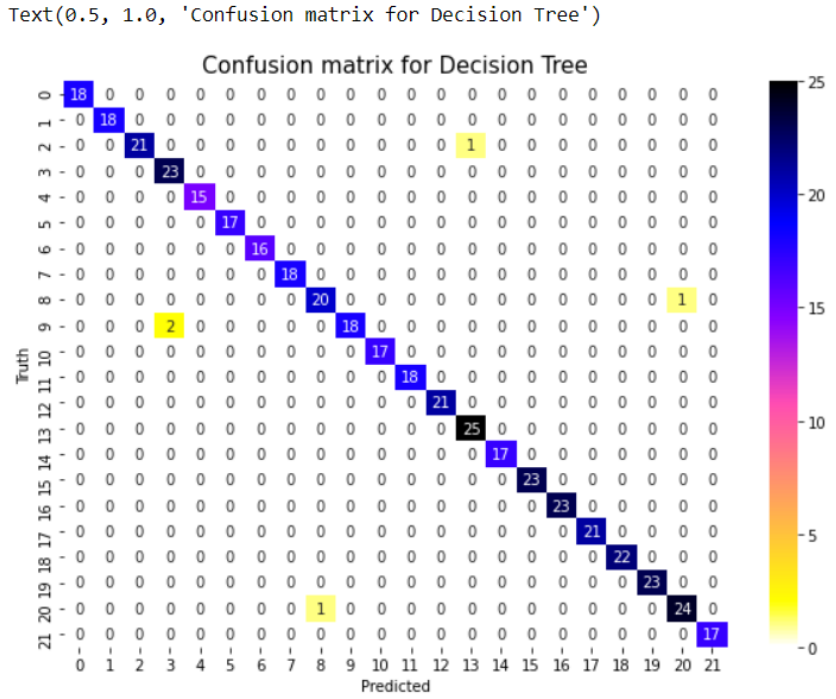

Figure 7 Confusion matrix for Decision Tree

Figure 8 shows a Graphical Representation of the Accuracy Score of various Classification Algorithms.

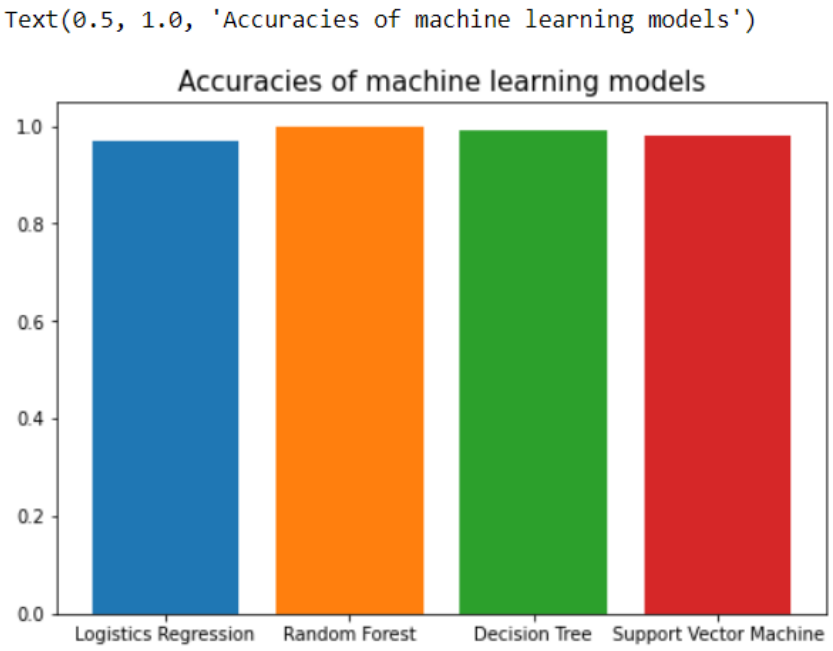

Figure 8 Bar graph for Result Analysis

5. Conclusion

Once this work is completed, we will be able to predict the plants according to the soil and the recommended climate. The proposed plan lists all potential crops for a particular region, helping the farmer to decide which crop to plant. The program has conducted a comprehensive assessment of soil, climate and pH knowledge and shows which plants are most beneficial to be grown in the right environment. Direct agriculture is still desirable in many developing countries. It is possible to achieve the above vision in India to improve food security and the individual income of farmers. The above-mentioned challenges and promising solutions to predict the future of Indian agriculture. Technological advances and the government's efforts to promote and promote agriculture with the right resources, assistance, tax holidays and other benefits to farmers will greatly attract investment. This initiative will therefore help deliberate efforts to prevent the growth and sustainability of future generations.

References

1. Ashok T., Suresh Varma P. (2020). Crop Prediction Based on Environmental Factors Using Machine Learning Ensemble Algorithms. In: Peng SL., Son L., Suseendran G., Balaganesh D. (eds) Intelligent Computing and Innovation on Data Science. Lecture Notes in Networks and Systems, vol 118. Springer, Singapore. https://doi.org/10.1007/978-981-15-3284-9_67.
2. V. Saiz-Rubio and F. Rovira-Más (2020). “From Smart Farming towards Agriculture 5.0: A Review on Crop Data Management,” Agronomy, vol. 10, no. 2, p. 207, Feb. 2020.
3. Victor Mokaya (2019). “Future of Precision Agriculture in India using Machine learning and Artificial Intelligence,” International Journal of Computer Sciences and Engineering, Vol.7, Issue.3, pp.422-425, 2019.
4. Sharma, A. Jain, P. Gupta and V. Chowdary (2021). "Machine Learning Applications for Precision Agriculture: A Comprehensive Review," in IEEE Access, vol. 9, pp. 4843-4873, 2021, DOI: 10.1109/ACCESS.2020.3048415.
5. Mupangwa, W., Chipindu, L., Nyagumbo, I. et. al., (2020). Evaluating machine learning algorithms for predicting maize yield under conservation agriculture in Eastern and Southern Africa. SN Appl. Sci. **2,** 952 (2020). https://doi.org/10.1007/s42452-020-2711-6.
6. D. Elavarasan, D. R. Vincent P M, K. Srinivasan, and C.-Y. Chang (2020). “A Hybrid CFS Filter and RF-RFE Wrapper-Based Feature Extraction for Enhanced Agricultural Crop Yield Prediction Modeling,” Agriculture, vol. 10, no. 9, p. 400, Sep. 2020.
7. S. Saranya and T. Amudha (2016). "Crop planning optimization research — A detailed investigation," 2016 IEEE International Conference on Advances in Computer Applications (ICACA), 2016, pp. 202-208, DOI: 10.1109/ICACA.2016.7887951.
8. S. A. Bhat and N. -F. Huang (2021). "Big Data and AI Revolution in Precision Agriculture: Survey and Challenges," in IEEE Access, vol. 9, pp. 110209-110222, 2021, DOI: 10.1109/ACCESS.2021.3102227.
9. Thomas van Klompenburg, Ayalew Kassahun, Cagatay Catal (2020). “Crop yield prediction using machine learning: A systematic literature review,” Computers and Electronics in Agriculture, Volume 177, 2020, 105709, ISSN 0168-1699, https://doi.org/10.1016/j.compag.2020.105709.
10. Shahhosseini, M., Hu, G., Huber, I. et. al., (2021). Coupling machine learning and crop modelling improve crop yield prediction in the US Corn Belt. Sci Rep **11,** 1606 (2021). https://doi.org/10.1038/s41598-020-80820-1.

11. K. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis (2018). "Machine Learning in Agriculture: A Review," Sensors, vol. 18, no. 8, p. 2674, Aug. 2018 [Online]. Available: http://dx.doi.org/10.3390/s18082674.
12. Balafoutis, Athanasios & Beck, Bert & Fountas, Spyros & Vangeyte, Jürgen & van der Wal, Tamme & Soto, Iria & Gómez-Barbero, Manuel & Barnes, Andrew & Eory, Vera. (2017). Precision Agriculture Technologies Positively Contributing to GHG Emissions Mitigation, Farm Productivity and Economics. Sustainability. 9. 1339. 10.3390/su9081339.

***