International Journal of Scientific Research & Engineering Trends
Volume 8, Issue 3, May-Jun-2022, ISSN (Online): 2395-566X

Agricultural Crop Recommendations System Based on Soil Fertility Using Machine Learning Techniques

Sneha A.S, Nikila A, M. Dharun Gandhi Supervisor: Mr.P.Boopathirajan
Department of computer Science Engineering,
Dr.Mahalingam College of Engineering and Technology
TN,India

Abstract- Agriculture plays a major role in everyday life, without proper agriculture techniques production of quality crops will be affected. This may lead to economic crisis or downfall. Agriculture gives a significant hand in country's economic development. It is mandatory to administer new farming techniques and include technological advancements for better crop yield and production. Machine Learning techniques have greatly shaped agriculture throughout time. Machine Learning techniques greatly help in recommendation and prediction systems. Soil techniques greatly help in recommendation and prediction systems. Soil fertility is vital for healthy and increased crop yield. Soil fertility is determined based on several factors. Some of the features used in this approach are Nitrogen (N), Phosphorus (P), Potassium (K), humidity, temperature, rainfall and ph value of the soil. In our approach we use the technique of Multilayer Perceptron Classifier and Stochastic Gradient Descent to predict the best crop yielded based on the soil fertility.

Keywords- Crop Recommendation, Soil Fertility, Machine Learning, Gaussian Naïve Bayes, Decision Tree, Stochastic Gradient descent Classifier, Multilayer Perceptron.

I. INTRODUCTION

Agriculture plays a very important role in the supportability of a country. India is one of the countries that can yield a wide variety of agricultural products. One of the agricultural products that hold important products is horticulture products such as rice, corn, soybean, etc. Maintaining and increasing food security in a country is mandatory in building the economy.

Agricultural problems that occur in India are very crucial depending on the location of an area. Innovation technological ideas should be administered in farming techniques, to increase the productivity of agricultural products both in terms of quality and quantity. With massive population like India, agricultural products are in high demand. Hence increased production and good quality crops should be cultivated.

The use of agricultural land that has shifted function, the difficulty of expansion of agriculture land, control of population growth, and poor agricultural yields due to seasonality and prolonged weather are some of the problems that faced in one solution to improve agriculture yield is to adjust the plants to be planted with factors that affect the growth of food crops such as rainfall, temperature, altitude, and pH. Soil fertility is the major factor in increased and quality crop production.

In this paper we have applied different machine learning algorithms to find out the prediction accuracy based on the historic results of one of the Crop. These models are useful for predicting the crop yield. If the model predicts a poor crop or shows if a particular crop has the probability to fail in upcoming yield then respective algorithm strategies would be employed to make the crop improved.

Four machine learning techniques Gaussian naïve bayes, Decision tree, Multilayer perceptron classifier, Stochastic Gradient Descent classifier algorithms are used to build the machine learning model. ROC index has been used to compare the accuracy of the models. The dataset which have been used to build the model was collected from one of the agricultural crop across the year 2017, 2018, 2019 and 2020. The dataset contains 950 entries and 8 attributes on that 200 entries are trained it contains 9 types of crops. The data contains Nitrogen (N), Phosphorus (P), Potassium (K), Temperature, PH value of soil, Rainfall, Humidity as an attributes and crop recommendation as a target.

II. METHODOLOGY

1. Gaussian Naïve Bayes:

Naïve bayes algorithm is used for large amount of information and also it is easy and rapid classification method. Gaussian Naive bayes support the continuous value for features and models. This classifier is employed when the predict values from continuous and they are expected to follow a Gaussian distribution. It is a supervised learning algorithm. Its assumption is that each class follows a Gaussian distribution. Naive bayes contains class specific covariance matrices.

2. Decision Tree:
Decision tree algorithms are used to solve both the classification and regression methods. It is used as a tree representation to solve the problem. Each leaf nodes in the tree correspond to the class labels and the attributes are represented for the internal node. Decision tree is a graphical representation to get all possible solutions for a problem or statements based on given condition.

In this algorithm, to predict the crops in the dataset the algorithm starts from the root node of the tree. It compares the values of root node attributes with the actual dataset or record attributes based on the comparison, it follow the branch and transfer to the next node. Again the algorithm compares with the attribute values to the other sub nodes and it moves further. The process will be done once it reaches the leaf node of the tree. Shown in Fig 2

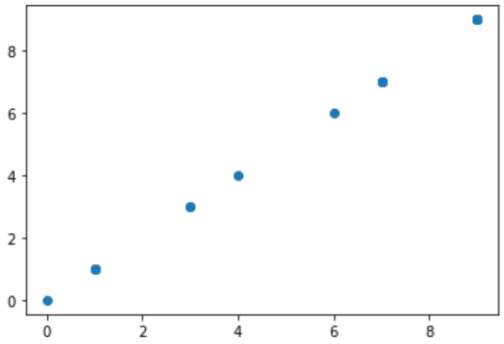

Fig 2. How decision tree works.

3. Stochastic Gradient Descent:
The SGD works with the random probability, for each iteration a few samples are randomly collected from the data set instead of the whole data. The word batch is known as the total number of samples from a data set which is used for calculating the gradient descent for each iteration. In SGD, it uses only a single sample that is a batch of one to perform each iteration. The sample is randomly selected for the iteration.

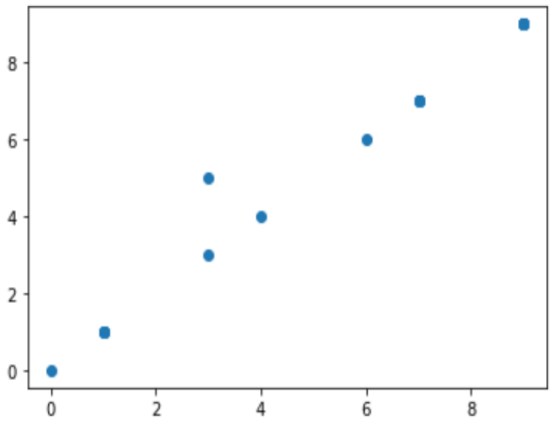

Fig 3. How stochastic gradient descent works.

In SGD we can find the cost function for the single sample at each iteration instead of the sum of the gradient of all the samples, since only one sample is chosen at random for dataset for each iteration, the path taken by the algorithm to reach the minimum is usually noisier than the gradient descent algorithm. But that's does no matter because all the path taken by the algorithm as longer as we reach the minima and with the significant shorter time.

SGD requires the higher number of iteration compared to typical gradient descent to reaches the minima, it is still computationally less expensive. So, SGD is preferred for optimizing the learning algorithm, shown in Fig 3.

4. Multilayer Perceptron:
Multilayer perceptron is a fully connected feedforward artificial neural network, MLP is a classical type of neural networks, It has an input and output layers and one or more hidden layers with heavy neurons are linked together, it is used to research for their ability to solve problems. MLP is fully connected with each node in one layer connects with the certain weight to every node in the layer.

III. EXPERIMENTATION

1. Data Collection:
The dataset used in this research is collected from one of the agricultural crops dataset from the years between 2017 and 2020. The selected dataset contains 950 rows of data and an approximate of 80% of the data is used for training and 20% of the data is used for testing. The prediction program uses the fields Nitrogen (N), Phosphorus (P), Potassium(K), Temperature, PH value of soil, Rainfall, Humidity as input attributes and the crop recommendation is provided as output or target.

2. Data pre processing and Machine Learning Software:
By using range query, the mean of each of the fields with numeric value has been found and to reduce the time to compute the model, the mean value found has been subtracted from each of the row attribute entries. Hence, it effectively reduces the number of numerical bits to process and therefore, the data is normalized. Machine learning models present in the python scikit and tensorflow libraries have been utilized for the purpose of testing, training of data and to infer the results.

3. Validation Method:
Three-fold cross validation has been done in this research. The dataset has been divided into 80% to 20% ratio where, 80% of data is taken for training and 20% of data is for testing. Using the training data the ML model is trained and using the trained model, the test data is fed and the crop prediction of the test data is collected. For model validation, the predicted crop classes are compared along

with the already present target value in the test dataset and the correlation metrics are calculated. ROC index curve is used to measure the accuracy of the classification model. F-measure, precision and recall has also been calculated along with accuracy.

4. Pattern Identification:

With the results crop results obtained from the validation method, regular patterns in the dataset are identified. Using the identified pattern, the test input data are mapped and the calculated results are compared with the target test results to find the accuracy of the algorithmic models. Various methods are available for calculating the accuracy of the model using the recorded results namely Confusion matrix, etc. By implementing this method and comparing the accuracy scores of various algorithms namely Gaussian naïve bayes, Decision tree, Stochastic gradient descent, etc, Multilayer perceptron has been found to be the algorithms with best working accuracy and hence has been implemented for this purpose.

IV. RESULT

Four classification models has been implemented and tested using four machine learning techniques, Gaussian naïve Bayes, Decision Tree, Stochastic Gradient Descent, Multilayer perceptron. Multilayer perceptron gives better result of 96 percentage of accuracy as compared to Gaussian Naïve Bayes, Decision Tree and Stochastic Graadient Descent, Decision Tree showed 60 percentage of accuracy, Gaussian Naïve Bayes showed 68 percentage of accuracy and Stochastic Gradient Descent showed 95 percentage of accuracy.

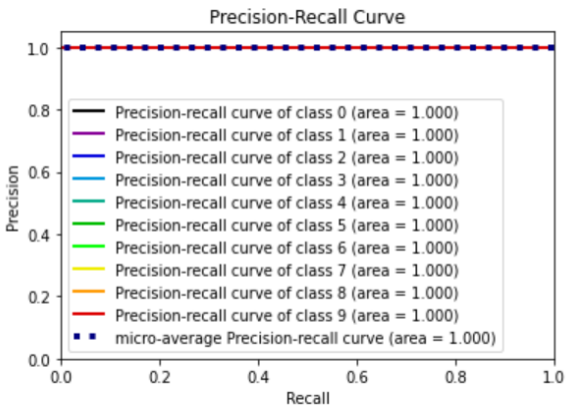

Fig 4. ROC index for Multilayer perceptron.

V. CONCLUSION

Stochastic gradient descent and Multilayer preceptor can be used to predict the crops for the particular soil, splitting the data into training and testing models into 4:1 ratio gives more accurate results. The training data set performed well during the prediction with an accuracy of 95 percentage as SGD and 96 percentage as MLP.

Moreover, model based on Multilayer Perceptron was trained to predict the crop outcome in the particular soil which predicts records with 96% accuracy. The Multilayer perceptron predicts the crops accurately while giving the proper inputs. However, it does not predict well when we have entered the improper values.

REFERENCES

[1] Shreya S. Bhanose, Kalyani A. Bogawar (2016) “Crop And Yield Prediction Model”, International Journal of Advance Scientific Research and Engineering Trends, Volume 1,Issue 1, April 2016.

[2] Tripathy, A. K.,"Data mining and wireless sensor network for agriculture pest/disease predictions." Information and Communication Technologies (WICT), 2011 World Congress on. IEEE at 2011.

[3] Ramesh Babu Palepu An Analysis of Agricultural Soils by using Data Mining Techniques”, International Journal of Engineering Science and Computing, Volume 7 Issue No. 10 October at 2017.

[4] Rajeswari and K. Arunesh “Analysing Soil Data using Data Mining Classification Techniques”, Indian Journal of Science and Technology, Volume 9, May 2016.

[5] Pritam Bose, Nikola K. Kasabov, “Spiking Neural Networks for Crop Yield Estimation Based on Spatiotemporal Analysis of Image Time Series”, IEEE Transactions On Geoscience And Remote Sensing at 2019.

[6] Priyanka P.Chandak ,” Smart Farming System Using Data Mining”, International Journal of Applied Engineering Research, Volume 12, Number 11 at 2017.

[7] Nasrin Fathima.G , “Agriculture Crop Pattern Using Data Mining Techniques”, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, May at 2014.

[8] Ramesh A.Medar, ”A Survey on Data Mining Techniques for Crop Yield Prediction”, International Journal of Advance Research in Computer Science and Management Studies, Volume 2, Issue 9, September at 2019.

[9] Savae Latu , ”Sustainable Development : The Role Of Gis And Visualisation”, The Electronic Journal on Information Systems in Developing Countries, at 2017 EJISDC 38, 5, 1-17.

[10] Konstantinos G. Liakos , Patrizia Busato , Dimitrios Moshou , Simon Pearson ID and Dionysis Bochtis “Machine Learning in Agriculture”Institute for Bio-Economy and Agri-Technology (IBO), CERTH Received: 27 June 2018; Accepted: 7 August 2018; Published: 14 August 2018.

[11] Prof. D.S. Zingade, Omkar Buchade, Nilesh Mehta, Shubham Ghodekar, Chandan Mehta “Crop Prediction System using Machine Learning” International Journal of Advance Engineering and Research Development Special Issue on Recent

Trends in Data Engineering Volume 4, Special Issue 5, Dec.-2017.

[12] Shakil Ahamed.A.T.M, Navid Tanzeem Mahmood (2015)," Applying data mining techniques to predict annual yield of major crops and recommend planting different crops in different districts in Bangladesh", ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD),IEEE,June

[13] Vikas Kumar, Vishal Dave (2013), "KrishiMantra: Agricultural Recommendation System", Proceedings of the 3rd ACM Symposium on Computing for Development, January.