International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

COPY RIGHT

IJIEMR Transactions, online available on 27th Jun 2022. Link :http://www.ijiemr.org/downloads.php?vol=Volume-11&issue=ISSUE-06

DOI: 10.48047/IJIEMR/V11/I06/112

Title **CROP YIELD PREDICTION USING MACHINE LEARNING ALGORITHM**

Volume 11, Issue 06, Pages: 1707-1715

Paper Authors
Mrs. G Sri Lakshmi, N. Gowthami ,A. Harish, K. Sri Phanindra Sai Teja

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

CROP YIELD PREDICTION USING MACHINE LEARNING ALGORITHM

Mrs. G Sri Lakshmi, Assistant professor, MTech, Department of IT, sree.gpk@gmail.com
N. Gowthami ,BTech, Department of IT, nallamothugowthami5220@gmail.com
A. Harish, BTech, Department of IT, harishatluri25@gmail.com
K. Sri Phanindra Sai Teja , BTech, Department of IT, phanindra213@gmail.com

ABSTRACT: Agriculture is an important sector in India. It is the major source of economy. Agriculture is first and foremost factor which is important for survival. Machine learning (ML) could be a crucial perspective for acquiring real-world and operative solution for crop yield issue. Considering the present system including manual counting, climate smart pest management and satellite imagery, the result obtained aren't really accurate. So the researchers, scientists are trying their best to predict crop in a better way. This paper focuses mainly on predicting the yield of the crop by applying various machine learning techniques. In the proposed system the machine learning models are used and It uses the previous record for training the system using some machine learning algorithms like MLP and SVR where high accuracy can be achieved. Machine learning regression approach can be used for the better prediction of crop yield and improve the performance of the crop yield prediction.

1. INTRODUCTION

Crop production may be a complicated development that's influenced by soil and environmental condition input parameters. Agriculture input parameters vary from field to field and farmer to farmer. Collection such info on a bigger space may be a discouraging task. However, the environmental condition info collected in Republic of India at each 1sq.m space in numerous components of the district is tabulated by Indian meteoric Department. The massive such knowledge sets may be used for predicting their influence on major crops of that individual district or place. There are completely different foretelling methodologies developed and evaluated by the researchers everywhere the globe within the field of agriculture or associated sciences. A number of such studies are: Agricultural researchers in alternative countries have shown that tries of crop yield maximization through pro-pesticide state policies have LED to hazardously high chemical usage. These studies have reported a correlation between chemical usage and crop yield [1]. Agriculture is associate trade sector that's benefiting powerfully from the event of detector technology, knowledge science, and machine learning (ML) techniques within the latest years. These developments return to satisfy environmental and population pressures round-faced by our society, wherever reports indicate a requirement for robust international agriculture yield

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

increase to produce food for a growing population on a hotter planet.

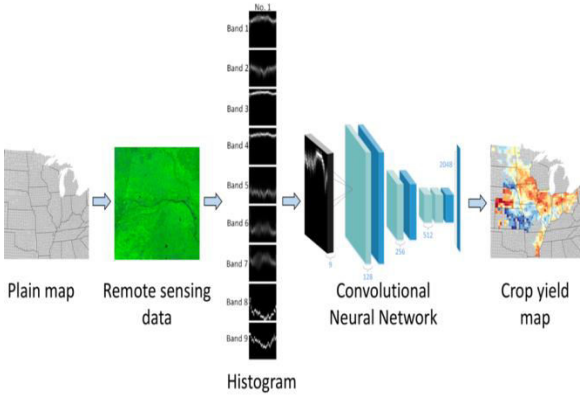

Fig.1: Crop yield analysis

Most of the work tired the sector of yield foretelling via cubic centimeter makes use of some kind of remote sensing knowledge over the farm. Agriculture seeks to extend and improve the crop yield and therefore the quality of the crops to sustain human. life. However, within the current time, folks tend to require a lot of like a shot appreciated jobs. There are fewer, and fewer folks concerned in crop cultivation. additionally, the continual increase of human population makes the cultivation of the crops at the proper time and right place even a lot of vital, because the climate is dynamic and therefore the shifts from traditional weather pattern are a lot of frequent than before manufacture. Food insecurity may be a drawback that can't be avoided, and humans should build use of latest innovative technologies to create use of existing soil, water and air conditions to get larger crops. The information gap between ancient ways that of cultivating and new agricultural technologies may be overcome if the computer code may be designed to model the interactive impact of climate factors, particularly the impact of maximum events (e.g. heat, rainfalls and excess water) occurring at completely different growing phases of crops. The temperature change undoubtedly affects the native and world food production, therefore planning computer code to model crop predictions needs new methodology for temperature change studies, situations for temperature change adaptation, and policymakers which will limit the devastating effects of weather on food provide. Experimental proof is employed to form environmental condition zones that have seen changes in weather and water, the 2 most significant factors in guaranteeing a in crop. The soil sort will modification over time because of weather and pests, therefore crop management must manage a fancy quantity of information, directly or indirectly associated with one another.

2. LITERATURE REVIEW

Agriculture Analysis Using Data Mining And Machine Learning Techniques

Agriculture is an important application in India. The modern technologies can change the situation of farmers and descision making in agricultural field in a better way. Python is used as a front end for analysing the agricultural data set. Jupyter Notebook is the data mining tool used to predict the crop production. The parameter includes in the dataset are precipitation, temperature, reference crop, evapotranspiration, area, production and yield for the season from January to December for the years 2000 to 2018. The data mining techniques like K-Means Clustering, KNN, SVM, and Bayesian network algorithm where high accuracy can be achieved.

Agricultural production output prediction using Supervised Machine Learning techniques

Farmers usually plan the cultivation process based on their previous experiences. Due to the lack of precise knowledge about cultivation, they end up cultivating undesirable crops. To help the farmers take decisions that can make their farming more efficient and profitable, the research tries to establish an intelligent information prediction analysis on farming in Bangladesh. However, this way of farming here is still at the initial stage. The research suggests area based beneficial crop rank before the cultivation process. It indicates the crops that are cost effective for cultivation for a particular area of land. To achieve these results, we are considering six major crops which are Aus rice, Aman rice, Boro rice, Potato, Jute and Wheat. The prediction is based on analyzing a static set of data using Supervised Machine Learning techniques. This static dataset contains previous years' data taken from the Yearbook of Agricultural Statistics and Bangladesh Agricultural Research Council of those crops according to the area. The research has an intent to use Decision Tree Learning-ID3 (Iterative Dichotomiser 3) and K-Nearest Neighbors Regression algorithms

Automated farming prediction

Farming is a predominantly manual process. The incorporation of any form of automation through the means of machine learning algorithms is still in incipient stage. This paper aims to introduce a fundamental approach to inaugurate the use of machine learning systems in the farming process. A comparative study between machine learning algorithms had been carried out in order to determine which algorithm is the most accurate in predicting the best crop for a particular land. Here, the best crop signifies the crop which had the most increase in terms of yield per unit area compared to previous years. This will ensure proper crop allocation throughout the country since optimal production for each crop will be secured. It will also increase the farmer's revenue margin. The study focuses on six major crops of Bangladesh which are Aman rice, Aus rice, Boro rice, Potato, Wheat and Jute. The algorithms that were used are Multiple Linear Regression (MLR) and K-Nearest Neighbor Regression (KNNR). Multiple Linear Regression (MLR) gave the most accurate results during the analysis and was incorporated into an android application. The android application system is also able to prepare a schedule of the complete farming process for a chosen crop, e.g. the correct time to apply fertilizers, irrigation, etc.

Improved Machine Learning Methodology for High Precision Agriculture

This paper presents the impact of machine learning in precision agriculture. State-of-the-art image recognition is applied to a dataset composed of high precision aerial pictures of vineyards. The study presents a comparison of an innovative machine learning methodology compared to a baseline used classically on vineyard and agricultural objects. The baseline uses color analysis and can discriminate interesting objects with an accuracy of (89.6%). The machine learning, an innovative approach for this type of use case, demonstrates that the results can be

improved to obtain 94.27% of accuracy. Machine Learning used to enrich and improve the detection of precise agricultural objects is also discussed in this study and opens new perspectives for the future of high precision agriculture.

A Scalable Machine Learning System for Pre-Season Agriculture Yield Forecast:

The system projected during this work is created by a neural network wherever inputs area unit treated on an individual basis. Static soil information in handled by fully-connected layers whereas dynamic meteorological information is handled by continual LSTM layers. This explicit design was trained with historical information for many soil properties, precipitation, minimum and most temperature against historical yield labels at county level. When training, the model was tested in an exceedingly separate information set and showed comparable results with existing yield prognostication ways that create use of in-depth remote sensing data. the most important lesson learnt from our experiments is that it's attainable get ascendable yield forecast as a result of the projected neural network model will notice and exploit redundant info each within the soil and within the weather information. To boot, the model might be able to learn AN implicit illustration of the cycles of the crops evaluated during this paper, considering the seasonal atmospherically information used as input.

Machine learning approach for forecasting crop yield based on climatic parameters

The present study provides the potential use of information mining techniques in predicting the crop yield supported the environmental condition input parameters. The developed webpage is user friendly and therefore the accuracy of predictions square measure higher than seventy-fiveper cent all told the crops and districts designated within the study indicating higher accuracy of prediction. The user-friendly web content developed for predicting crop yield may be utilized by any user their alternative of crop by providing environmental condition knowledge of that place.

Crop Prediction on the Region Belts of India: A Naïve Bayes MapReduce Precision Agricultural Model

The planned work introduces efficient degree economical crop recommendation system. Use of naïve mathematician makes the model terribly economical in terms of computation. The system is scalable because it may be wont to take a look at on totally different crops. From the yield graphs the simplest time of sowing, plant growth and gather of plant may be known. Conjointly the best and worst condition may also be incurred. The model focuses on all style of farms, and smaller farmers may also be benefitted.This model may be more increased to seek out the yield of each crop, and for chemical recommendation. Conjointly it may be changed to recommend concerning the fertilizers and irrigation want of crops.

3. IMPLEMENTATION

An agro-based country depends on agriculture for its economic growth. When a population of the country increases dependency on agriculture also increases and subsequent economic growth of the country is affected. In this situation, the crop yield rate plays a

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

significant role in the economic growth of the country. So, there is a need to increase crop yield rate. Some biological approaches (e.g. seed quality of the crop, crop hybridization, strong pesticides) and some chemical approaches (e.g. use of fertilizer, urea, potash) are carried out to solve this issue. In addition to these approaches, a crop sequencing technique is required to improve the net yield rate of the crop over the season. One of existing system we identified is Crop Selection Method (CSM) to achieve a net yield rate of crops over the season. We have taken example of CSM to demonstrate how it helps farmers in achieving more yield

Disadvantages:

- Accuracy is low**.**
- These segmentation have shortcomings
- Feature extraction is not accurate
- Accuracy will be low Computation load very high.

In proposed work, Support Vector Machine Support vector machine used in crop yield prediction is called support vector regression. The aim of the support vector technique is to obtain non-linear function using kernel function (a linear function or polynomial function). The radial basis function and the polynomial function are the widely used kernel function. The merit of support vector regression is to avoid difficulties of using linear function in large input samples space and optimization of a complex problems transformed into simple linear function optimization.

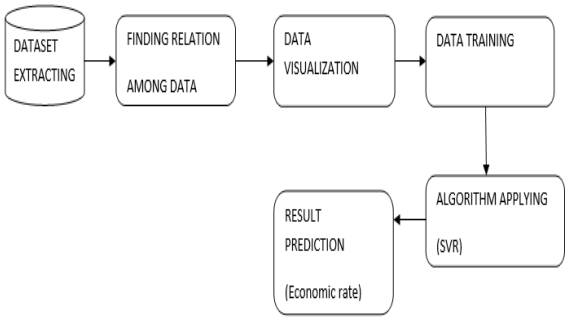

Fig.2: System architecture

In our project the crop yield classification will perform to categorize on the basis of yield productivity and class labels will be in range. And range of productivity will be defined and regression will be performed to get the actual crop yield estimated range. This is the motive to develop this system. Based on crop weather studies, crop yield forecast models are prepared for estimating yield much before actual harvest of the crops. By use of empirical statistical models using correlation and regression technique crops yield are forecast on an operational basis for the country. Meteorological parameters at various crop growth stages along with technological trends are used in the models. And this research will also helpful if in future we make a complete recommender system for farmers. Because here we are performing descriptive analytics which is the base or foundation of any recommender system.

The dataset that is used needs to be pre-processed because of the presence of redundant attributes, noisy data in it. Initially, data cleaning operation is performed where the redundant factors are determined and are not considered for the prediction of crops. Over18 which are either having the same

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

values for all the crops are completely unrelated to the prediction task. As part of the exploratory data analysis, the categorical factors are split and are assigned values as 0 and 1 based on whether the factor is present or not. These assigned values assist in further classification based on that particular factor.

Advantages:

- Speed and very low complexity, which makes it very well suited to operate on real scenarios.
- Computation load needed for image processing purpose is much reduced, combined with very simple classifiers..
- Ability to learn and extract complex image features.
- With its simplicity and fast processing time, the proposed algorithm is suitable to be implemented in embedded system or mobile application that has limited processing resources

4. ALGORITHM

Supporting vector Regression (SVR):

A Support Vector Regression (SVR)is a discriminative classifier formally defined by a separating hyperplane. In other words, given labeled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorizes new examples. In two dimensional space this hyperplane is a line dividing a plane in two parts where in each class lay in either side.

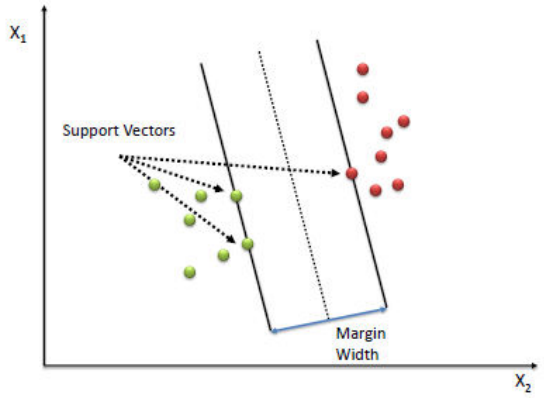

Fig.3: SVM model

SVM (Support Vector machine) is a supervised learning algorithm, and is one of the most efficient and universal classification algorithms. Its goal is to fmd the optimal separating hyperplane which maximizes the margin of training data. Initially the classifier is trained with labelled data before being used to classify the data to test accuracy. Before the data can be used to train our classifier, it is imperative to process it. This consists of the following steps:

- Labelling of data
- Generation of vocabulary
- Creation of document-term matrix

Multilayer Perceptron:

A multilayer perceptron can be defined as a class of feedforward artificial neural networks (ANNs). Multilayer perceptrons are sometimes referred to as neural networks, especially when they have a single hidden layer. A Multi-Layer Perceptron or MultiLayer Neural Network contains one or more hidden layers (apart from one input and one output

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

layer). While a single layer perceptron can only learn linear functions, a multilayer perceptron can also learn non – linear functions. An MLP consists of at least three significant layers, an input layer, a hidden layer, and an output layer. Except for the input nodes, each node in the network is a neuron that makes use of a non-linear activation function. MLP utilizes a supervised learning technique called backpropagation for training. Its multiple layers and non-linear activation distinguish MLP from a linear perceptron. It can distinguish data that is not linearly separable.

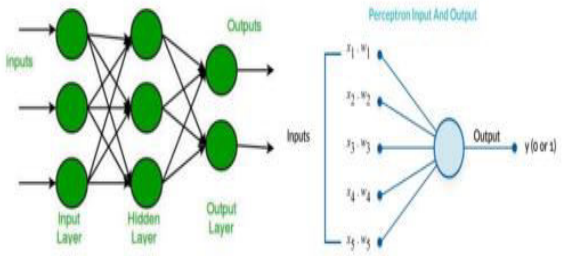

Fig.4: MLP model

5. EXPERIMENTAL RESULTS

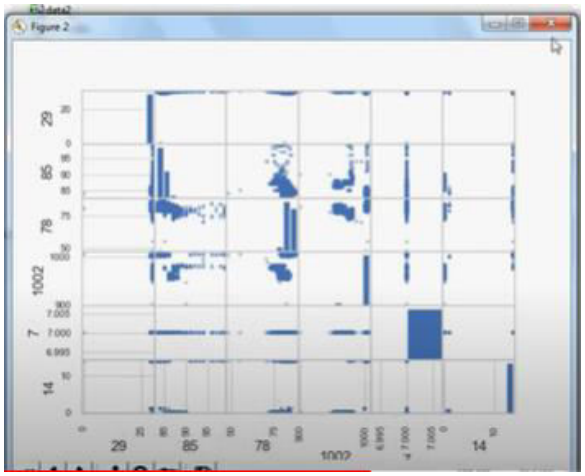

Fig.4: Output

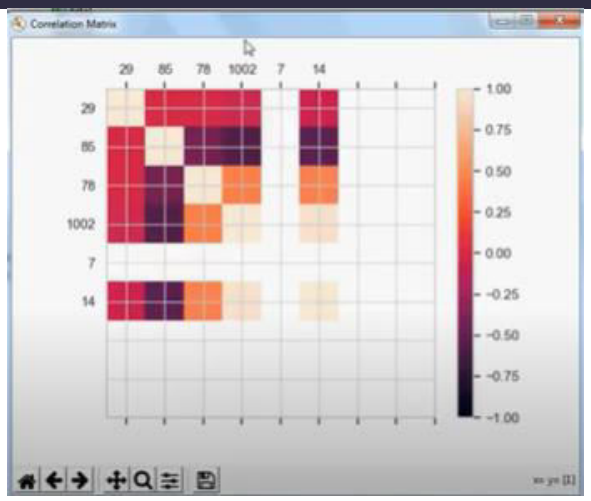

Fig.5: Output

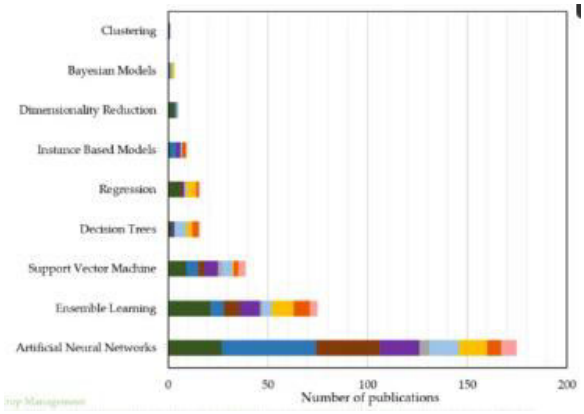

Fig.6: Output

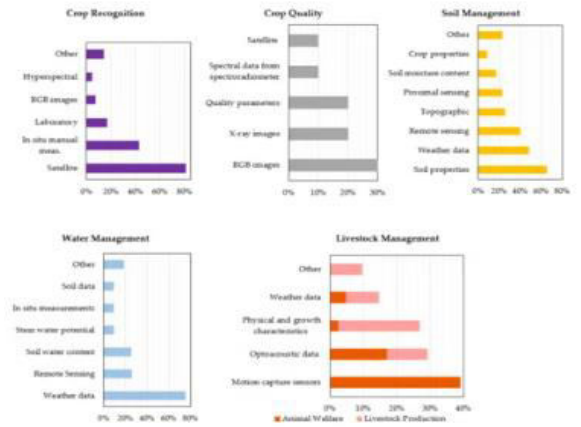

Fig.7: Output

6. CONCLUSION

The report includes the idea to implement the concept of descriptive analytics in agriculture domain. The research work provides the information about how could be apply data analytics on sugarcane crop datasets. There are three dataset named as Soil dataset, Rainfall dataset, Yield dataset. These datasets includes several parameters which are helpful to know the condition of crops and classify the data into separate classes by performing supervised training on the dataset that are collected from agriculture domain. This system has the capability to perform both the classification as well as regression. In the classification step the data is classified into three classes (low, mid, and high), whereas in regression step the actual cost of yield production is estimated. We used three major algorithms of supervised learning such as SVM and MLP to train and build a model. This work is basically provides the comparative study of various algorithm when we apply these algorithm on datasets and it shows the accuracy of each algorithms to train the datasets and also mean squared error at the cross-validation phase of the sample data. This work is domain independent. It means we can build system for other domain like as medical, product comparison, retails etc. We just need to pass the datasets through this system but dataset should be in consistent form. This research work can be enhancing to the next level. We can build a recommender system of agriculture production and distribution for farmer. By which farmers can make decision in which season which crop should sow so that they can get more benefit. This system is work for structured dataset. In future we can implement data independent system also. It means format of data whatever, our system should work with same efficiency.

7. FUTURE WORK

Sensor technologies are implemented in countless farming sectors. This paper aids in getting maximum yield rate of the crops. Also assists in selecting proper crop for their selected land and selected season. These techniques will resolve the problems of farmers in agriculture field. This will aid in improving the economic growth of our country. Along with that the rainfall prediction is also done to estimate theuield of the crop based upon the rainfall in the land.

REFERENCES

[1] Veenadhari S, Misra B, Singh CD. Data mining techniques for predicting crop productivity—A review article. In: IJCST. 2011; 2(1).

[2] Jain A, Murty MN, Flynn PJ. Data clustering: a review. ACM ComputSurv. 1999;31(3):264–323.

[3] Berkhin P. A survey of clustering data mining technique. In: Kogan J, Nicholas C, Teboulle M, editors. Grouping multidimensional data. Berlin: Springer; 2006. p. 25–72

[4] V. Arulkumar. "An Intelligent Technique for Uniquely Recognising Face and Finger Image Using Learning Vector Quantisation (LVQ)- based Template Key Generation," International Journal of Biomedical Engineering and Technology 26, no. 3/4 (February 2, 2018): 237-49.

[5] Han J, Kamber M. Data mining: concepts and techniques. Massachusetts: Morgan Kaufmann Publishers; 2001.

[6] C.V. Arulkumar, G. Selvayinayagam and J. Vasuki, "Enhancement in face recognition using PFS using Matlab," International Journal of Computer Science & Management Research, vol. 1(1), pp. 282-288, 2012

[7] H. Anandakumar and K. Umamaheswari, "A bio-inspired swarm intelligence technique for social aware cognitive radio handovers," Computers & Electrical Engineering, vol. 71, pp. 925–937, Oct. 2018.

[8] Ester M, Kriegel HP, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. 9