Crop Prediction using Machine Learning

N.L. Chourasiya, P. Modi , N. Shaikh[3], D. Khandagale, S. Pawar

(Department of Computer Engineering, MES College of Engineering/S.P.Pune University, India)

Abstract: *In Indian history, agriculture has been the backbone of the economy. This agricultural activity stay undeveloped thanks to various factors. Most of the activities area unit through with an absence of recent technology. Currently, seed classification is completed supported information of person. the aim of this study is to look at the prediction of crops which will offer high yield within the given location considering the climate and soil parameters.*

The current seed classification analysis is inefficient and has no validation mechanism. during this analysis, we've created an attempt to gift a prophetic model to predict seed sowing for farmers mistreatment machine learning algorithmic program which ends up in high crop production. For the event, this analysis machine learning algorithmic program is employed to be told from information which might be wont to build predictions, to create real-world like simulations, for pattern recognitions and classifications of the input file. a man-made neural network is employed for modeling advanced relationships between inputs and outputs or to search out patterns in information.

The objective of this thesis is to grasp the machine learning algorithmic rule mistreatment neural networks and constructing model that predicts seed categories supported machine learning technique. The model is experimented mistreatment seed dataset and so seed categories area unit foreseen mistreatment the developed model.

Keywords: *Crop Prediction, Climate Factors, Soil Factors, Study of ML techniques for Prediction, ANN, Multi-linear regression, SVM .*

I.   Introduction

The usage of technology in agriculture works has started since the first twentieth century once the business enraptured from the horse drawn digger to mechanized tractors. The introduction of plant heredities, chemical inputs and crop management systems has reworked the business into technology enabled and data-rich world. The technological progress that compose this computing setting have contributed to discussing concerning huge information whereas information collection isn't new ideas particularly within the context of public information assortment. the sole begin of a lot of economical mobile technologies and therefore the conversion of knowledge have allowed giant records be evaluated and analyzed in a very timely and a lot of helpful ways that.

Agriculture is that the most significant economic sector of the many developing countries. Indian agricultural activities have continued underdeveloped ways in which due to lack of getting ample technologies. the opposite reason behind unproductivity is drought, that has often affected the countries agricultural activities since the first Seventies. This downside leads low productivity, weak infrastructure, low level of technology and overspill. because of this multitude issues, Indian farmers would like skilled recommendation to own a lot of productive.

There is a requirement for economical framework to anticipate and enhance the yield.[1] so as to require full advantage of the soil kind, moisture, humidity, climate and etc. farmers have to be compelled to grasp precisely the form of seeds for his or her cropping. completely different districts in Asian country have varied climates then it's vital to think about environmental factors of those separate areas. This helps to decide on the most effective districts for farming of various form of seeds. precipitation conjointly varies from district to district and this features a immense impact on farming as a result of whereas deficient or an excessive amount of rain will kill crops, the right quantity of rain results in good crop yield. In today's conditions, agricultural enterprises area unit capable of generating giant amounts of knowledge. therefore this Growth in knowledge size needs an automatic technique to extract and analysis necessary knowledge. Machine learning algorithms, ANN area unit wont to support agricultural center consultants. ANN holds one among the keys for farmers management centers to gather and method knowledge in real time to assist farmers that creates the most effective choices with relation to planting, fertilizing and harvest crops. In today's conditions, agricultural enterprises area unit capable of generating and aggregation giant amounts of knowledge. therefore this Growth in knowledge size needs an automatic technique to extract and analysis necessary knowledge.

We projected associate degree automatic seed category predictor model that classifies seed dataset mistreatment ANN machine learning tool. An ANN, may be a kind of computer science that consists of an

outsized range of straightforward process elements known as artificial neurons or nodes that ar interconnected by direct links, known as connections, and that collaborate to perform parallel distributed process (PDP) operation so as to unravel a given drawback. A subgroup of process element is named a layer within the network. the bottom layer is that the input layer and also the highest layer is that the output layer. Between the bottom and highest layer, there could also be a further layer(s) of units, known as hidden layer(s). The advantage of neural networks over typical programming lies in their capability to become answer|an associate degreeswer} for various issues that don't have an algorithmic resolution or the accessible solution is simply too advanced to be found. associate degree ANN is adjusted for a particular application, like pattern recognition or knowledge classification, through a coaching method.

The ANN modeling is changing into very hip in numerous areas of agriculture, specially, within the areas wherever straight applied math modeling becomes unsuccessful. The ANN is victimization within the field of agriculture to predict the crop yield, biomass production, seeding dates, physical and physiological damaging of seeds, organic matter contents within the soils, soil wet estimation, mechanics properties of crops, estimation of sugar content in fruits and characterization of crop varieties.

In our analysis, we've thought-about the consequences of geometric parameters towards seed classification in Bharat. Taking these factors into thought as datasets for numerous districts, we have a tendency to applied appropriate model that predicts seed categories supported machine learning technique.

II. Project Scope

The objective of this study is to grasp the machine learning rule exploitation neural networks and constructing model that predicts seed categories supported machine learning technique. The model is experimented exploitation seed dataset and so seed categories ar foretold exploitation the developed model. The totally different parts like atmosphere, the type of soil and its arrangement, topography of the district, crop yield, market price have an effect on the selection of crop. We take into consideration, the climatic and soil factors. Machine learning gives numerous successful calculations which depends upon different factors.[2] It is a difficult task to identify the best suitable when there are more than one options available.[3] Hence, by machine learning, accurate crops can be predicted. We take into consideration, the climatic factors and soil attributes of location to predict the suitable crops.

III. Literature Review

3.1 Artificial Neural Network
Mahabadi et al. .[18]

In this paper, ANN models are ready with varied variety of neurons in hid layer, back propagation learning calculations. Modifying these parameters inspired the capability of the system to create up a perfect capability to foresee crop yield. the educational rate and variety of hidden nodes for the most part have an effect on show conduct. For the foremost half, less hidden nodes were needed because the quantity of knowledge diminished. the most effective models have less hidden nodes than the start variety of nodes. ANN models with a lot of nodes could have resulted in overfitting as hostile learning their connections. RMSE(Root mean sq. error), was used to assess the execution of created model.

Ranjeet et al.[19]

In this paper back propagation ANN is employed for foreseeing crop yield. variety of hidden layers, variety of neurons, area unit determined by conducting the trial and error technique iteratively. The neural system was ready with environmental condition and fertilizers use input. The environmental condition variables incorporate precipitation, Georgia home boy temperature and min temperature and fertilizers. total of squares is employed because the loss perform. Future work incorporates that cheap data parameters, correct range of hidden layers and range of neurons is chosen to boost the preciseness of forecast by utilizing the search algorithms, as an example, the genetic formula.

Meena et al. [20]

In this paper, back propagation ANN is used for harvest yield gauging. Twenty four factors area unit thought of because the data factors and therefore the harvest yield because the yield variable. The twenty four factors area unit traditional month to month precipitation, greatest temperature in a very month, least temperature in a very month, traditional month to month damp, mean ocean level weight, mean breeze speed (km/h), most extreme supported speed (km/h) and variety of long periods of pelter happened within the 3 months of harvest life cycle, as an example, July, August and Gregorian calendar month.

To evacuate repetition among data factors, affiliation investigation is employed. The projected technique acquires nice outcomes and truth of the current framework will be increased additional if

progressively physical components and different financial and innovative variables ar incorporated aboard harvest yield info for larger period.

Pandey et al. [21]

In this paper, 2 variants of ANN, the Radial Basis perform Neural Networks (RBFNN) and Generalized Regression Neural Networks (GRNN) for predicting potato crop yield. GRNN was a stronger robust|an improved} predictor than RBNN because it learns more quickly [21].

3.2 Multiple Linear Regression
Lefevre et al.

Regression analysis is employed to model the link between dependent and freelance variables. we tend to generate a perform that maps input to desired output. The coaching method of the model continues till the model achieves a desired level of accuracy on the coaching information. the overall equation of Multi regression is as follows:

Y is that the foretold outcome of this equation. The parameters, 0,1, 2 …. n ar the regression coefficients.123, n ar the freelance variables. during this case, climate and soil factors is thought-about as experimental variable and foretold crop as per the situation is variable.

Artificial Neural Networks ar higher than regression because it will acknowledge and exploit relationships within the information that aren't predefined (as in regression techniques) and want to not be speakable by any equation. This makes ANN significantly suited to map the relationships that are not-linear and empirical, on condition that ample information is obtainable to coach the network [22].

3.3 Support Vector Machine
Gandhi et al. [23]

This analysis has incontestable the prediction of rice crop yield by applying one among the machine learning technique, support vector machine (SVM). In terms of test's accuracy and quality conjointly Bayes internet and Multilayer Perceptron showed the very best accuracy and very best quality and SMO showed the bottom accuracy and worst quality. [23]

Vats et al. [24]

There square measure 3 dataset named as Soil dataset, rain dataset, Yield dataset. These datasets includes many parameters that square measure useful to grasp the condition of crops and classify the information into separate categories by playing supervised coaching on the dataset that square measure collected from agriculture domain. this method has the potential to perform each the classification further as regression. within the classification step the information is assessed into 3 categories (low, mid, and high), whereas in regression step the particular price of yield production is calculable. we have a tendency to used 3 major algorithms of supervised learning like KNN, SVM and LS-SVM to coach and build a model. This work is essentially provides the comparative study of varied formula once we apply these formula on knowledgesets and it shows the accuracy of every algorithms to coach the datasets and additionally mean square error at the cross-validation section of the sample data. This work is domain freelance. It suggests that we will build system for different domain like as medical, product comparison, retails etc. we have a tendency to simply have to be compelled to pass the datasets through this method however dataset ought to be in consistent type.

IV. Methodology And Algorithm
4.1 Data Acquisition and Filtration Algorithm
Input: Seed and Soil Data Set
Output: filtered information in key price combine and send this to process Mechanism
Steps:
1. Filter connected information i.e. Processed information. All alternative unessential information are going to be discarded.
2. Divide the info into acceptable Key price try.
3. Transmit unprocessed information on to aggregation step while not process.
4. Assign and transmit every distinct knowledge block of Processed knowledge to various process steps in processing Unit.

Description:
This formula takes Seed and Soil information Set and so filters and divides them into segments and performs approximation formula.

In step 1, connected details filtered out.

In step 2, filtered information ar the associated within the variety of totally different key price pairs. every combine of the particular key stores the various samples related to that key(types); which ends up in forming a knowledge block.

 In Next steps , these blocks ar forwarded to be processed by the info process Unit.

4.2 process and Calculation algorithmic program

Input: Filtered knowledge

Output: Normalized Result knowledge into Recommendations comparable kind in conjunction with Historical Values.

Steps:

1. for every event knowledge, relevant Historical knowledge is extracted.

2. Normalize this for all the information feed.

3. Persist {the knowledge|the info|the information} into data store and forward it.

V. Model

5.1 Artificial Neural Network

An ANN may be a assortment of connected units or nodes known as artificial neurons, that loosely model the neurons in brain. every association will transmit a its output to alternative artificial somatic cell as its input.

An artificial somatic cell that receives a input will method it so transmit output to the extra artificial neurons connected thereto. In common ANN implementations, the input of the synthetic neurons may be a complex number, and also the output of every artificial somatic cell is computed by some non-linear perform of the total of its inputs. Artificial neurons might have a threshold such the output is barely sent if it crosses that threshold.

Typically, artificial neurons square measure composed of various layers. These layers might perform totally different sorts of transformations on their inputs. The input is given to the primary layer, the output of every layer is fed to subsequent layers, and eventually to the last layer (the output layer), probably once traversing the layers multiple times.[15]

Most generally utilised topological calculation is multi-layered perceptron and back-propagation formula to implement neural networks prediction. sadly, there's no any automatic technique to work out an appropriate topology for knowledge sample house. Therefore, topology is by trial and error designated for appropriate crop yield prediction. a man-made neural network is employed once variety of input attributes is lesser.[16]

The general procedure to fabricate a neural network model incorporates the assembly of of information sets for coaching and testing, coaching multiple networks with totally different input parameters, result analysis, and model testing [17] .

VI. Conclusion

Various techniques via artificial neural network, multi-linear regression, and support vector machine are studied. It is concluded that ANN is a suitable method for the system, since the number of inputs is less. ANN is better than MLR in exploring non-linear relations between variables.   For checking the performance of the model, RMSE techniques can be used.

References

[1].   Utkarsha , P. Narkhede , K. P. Adhiya, valuation of Modified K-Means Clustering Algorithm in Crop Prediction, *International Journal of Advanced Computer Research , E Vol 4 , Issue 16,* September 2014

[2].   Karandeep Kaur, Machine Learning: Applications in Indian Agriculture , *International Journal of Advanced Research in Computer and Communication Engineering, Vol. 5*, Issue 4, April 2016

[3].   Mukesh Meena and Pramod Kumar Singh, Crop Yield Forecasting Using Neural Networks*, Springer International Publishing Switzerland* , pp. 319-331, 2013

[4].   Statistical Analysis of the Relationship Between Wind Speed, Pressure and Temperature*, Journal of Applied Sciences, volume 11, issue 15*, 2011

[5].   How do weather and climate influence cropping area and intensity? ,Global Food Security 4 (2015) 46-50

[6].   https://www.gardeningknowhow.com/plant-problems/environmental/temperature-on-plants.html

[7].   www.agritech.tnau.ac.in/agriculture/agri_agrometeorology_relativehumidity.html

[8].   A review of the effects of atmospheric humidity on the growth of horticultural crops, *Journal of Horticultural Science,* 62:2, 125-134, Nov 2015

[9].   Impact of monsoon rainfall on the total food grain yield over India, *J. Earth Syst.Sci. 123, No. 5, July 2014*, pp. 1129-1145

[10].   http://www.agr.gc.ca/eng/science-and-innovation/agricultural-practices/soil-and-land/soil-and-water/soil-texture-and-water-uality/?id=1197483793077

[11].   https://www.clearias.com/soils-of-india-classification-characteristics

[12].   https://homeguides.sfgate.com/list-crops-grow-sandy-soil-94331.html

[13]. https://www.maximumyield.com/definition/473/drainage
[14]. https://articles.extension.org/pages/63501/solutions-to-soil-problems:-soil-acidity
[15]. https://en.wikipedia.org/wiki/Artificial_neural_network
[16]. Rakesh Kumar , M.P. Singh , Prabhat Kumar and J.P. Singh, Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique*, International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials, .* pp. 138-145, may 2015
[17]. B Roner, I. & Comstock, C. R. (1997). Combining expert systems and neural networks for learning site-specific conditions. Computers and Electronics in Agriculture 19,37-53
[18]. Use of the Intelligent Models to Predict the Rice Potential Production, *International Academic Journal of Innovative Research, Vol. 2, No. 10, 2015*, pp. 20-31.
[19]. An Artificial Neural Network for Predicting Crops Yield in Nepal*, Asian Federation for Information Technology in Agriculture*, pp. 371-381
[20]. Talha Siddique, Dipro Barua, Zannatul Ferdous, Amitabha Chakrabarty, Automated Farming Prediction, *Intelligent Systems Conference 2017, 7-8 September 2017*
[21]. Application of Artificial Neural Networks in Yield Prediction of Potato Crop*, RUSSIAN AGRICULTURAL SCIENCES, Vol. 43, No. 3, 2017*
[22]. NATHALIE LEFEVRE ANDREW J. WATSON and ADAM R. WATSON, A comparison of multiple regression and neural network techniques for mapping in situ pCO2 data, University of East Anglia, *School of Environmental Sciences, Norwich NR4 7TJ, UK*
[23]. Rice Crop Yield Prediction in India using Support Vector Machines, *IEEE 2016*