IJBPAS, November, Special Issue, 2021, 10(11): 563-573

ISSN: 2277–4998

International Journal of Biology, Pharmacy and Allied Sciences (IJBPAS)

'A Bridge Between Laboratory and Reader'

www.ijbpas.com

ESTIMATION OF AGRICULTURAL PRODUCTION USING CATEGORIZATION METHODS IN THE CONTEXT OF BIG DATA ANALYSIS

SAKTHIVEL S[1*], CHIRRA KESAVA REDDY[2], KANNADASAN B[3], P. VISWANATH [4], JYOTHI N.M[5] AND KARPAGAM.M[6]

1: Professor in Computer Science and Engineering at Sona College of Technology, Salem, Tamilnadu

2: Professor& Principal in Mechanical Engineering at Universal College of Engineeing & Technology, Perecherla Andhra Pradesh, India

3: Assistant Professor, Civil Engineering, B.S.Abdur Rahman Crescent Institute of Science and Technology, GST Road, Vandalur Chennai – 600048

4: Assistant Professor (A), School of Management Studies, JNTUA, Ananthauramu- 515002, Andhra Pradesh

5: Associate Professor in Computer Science & Engineering at Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andra Pradesh, India

6: Professor in Electronics and Communication Engineering at Sri Krishna College of Engineering and Technology, Kuniamuthur, Coimbatore, Tamil Nadu, India

***Corresponding Author: Sakthivel S; E Mail: sakvel75@gmail.com**

Received 20th July 2021; Revised 22nd Aug. 2021; Accepted 30th Sept. 2021; Available online 1st Nov. 2021

https://doi.org/10.31032/IJBPAS/2021/10.11.1049

ABSTRACT

Agricultural is a primary generator for money in our nation and the foundation of our wealth. Agricultural surveillance can be available to landlords in order can help with existing issues including as flooding, uncontrolled spending due to requirements disparities, and climate instability. Agricultural productivity rates due to unpredictably changing climatic conditions, insufficient sanitary skills, soil water deterioration, and traditional agricultural viewpoints are all addressed. Augmentation training was another type of technology employed in agricultural to assess crop yield.Various reinforced teaching methods, such as predications, categorization, verification, and assemblages, are exploited to predict agricultural yield. Convolution neuronal systems, multilayered instructors, sequencing and regressive, predicting forests, and Multivariate Bayesian are some of the techniques used to include prognosis. Our investigators, on the contrary side, have a challenge in choosing the best approach among among the available options to those commodities they've defined. One major goal of this study is to see how agricultural output may be predicted using optimizing techniques.A strategy for estimating agricultural output employing categorization approaches has been described in the context of big information analysis.

Keywords: ISTA, IISTA, picture acquisitions, proportionate difficulties, Regularization functions, l0 standard, 11 standards, 12 information faithfulness phrase

INTRODUCTION

Cultivation, which comprises the production of crops, is a major contributor to the Indian economy. Harvests can be either edible or promotional in purpose. Paddy, wheat, cornmeal, millets, and other food sources exist alongside crop production such as sugarcane, jute, peanuts, and cashew. Weather conditions have a substantial influence on agricultural production **[1]**. As a consequence, reliable yield prediction is a substantial issue that must be addressed.Growers would be able to take sensible precautions to enhance productivity if yields could be anticipated ahead of schedule.Early forecasting can be accomplished by aggregating prior farmer experience, weather conditions, and other contributing attributes and storing them in a massive database **[2]**.

Determining agricultural output before cultivation is a critical issue in farming, as crop yield fluctuations affect international trade, food availability, and worldwide market pricing **[3]**. Preliminary crop production forecast also provides vital information to regulators.Agriculture production must be typically estimated to plan land use and monetary reforms adequately. Agricultural production monitoring on a field-by-field basis has become more common in recent years.Meteorological conditions have a significant impact on crop productivity. Growers can be warned well in preparation if temperature predictions are made more precise, which will assist to avoid large damages and accelerate economic growth **[4]**. In the event of a catastrophic emergency, the prognostication will also assist farmers in making decisions such as choosing alternative crops or discarding a crop at a preliminary phase.

Machine Learning is used to solve situations where the relationship between the input and output is unknown. The automated acquisition of architectural description was referred to as learning.Reinforcement learning, unlike traditional algorithms, does not decide the precise architecture of the analytical model that describes the evidence.It element comes in handy for modeling non-linear behavior like crop prediction. Intelligence is a computer technology **[5]** application that is used to produce efficient systems.The test specimens can be detected using the training samples.Metrics such as mean square error, precision, retention, susceptibility selectivity, and others can be used to assess the system's reliability.Additionally, supervised, unmonitored, and reinforcement learning methods could be used to address different domains, such as crop production **[6]**. Some of the methodologies used to build an adaptive approach include segmentation, clustering, regression, and forecast. Prediction was considered for the study, and the strategies for estimation are discussed in the next subsection.

Related Works

Agricultural development is affected by weather parameters such as moisture, temperature, soil quantity, terrain, or cultural aspects to the tune of 30% **[7]**. Probabilistic methods, extrapolation, convolution neural networks, data mining, support vector machines, and other approaches have been presented in the literature for determining crop production **[8]**. To anticipate the yield, an intelligent system like fuzzy logic uses logical reasoning.Nonetheless, obtaining the rules for forecasting necessitates frequent interactions with specialists.These criteria are also related to a specific set of input data **[9]**.

Numerous articles explore quadratic regression-based forecasting. It's a statistical technique for engaging with stability analysis.This method could be used to assess the relationship between the explanatory repressors.Multiple regressions could be used if the predictor variables have more than one input attribute.In prognostication, regression-based models are used when they produce consistent results in standard tests **[10]**. Although regression-based models perform well with models of computation, they are inefficient with complicated and non-linear data.Moreover, due to extrapolation assumptions for numerous co-linearity among the predictor variables, some models might not be able to outperform the competition.Tea Crop Yield was forecasted using a prediction model **[11]**. The majority of the observed events were properly anticipated, according to the findings.Estimation was presented for forecasting tea yield based on climate variables.An investigation of agricultural forecasting using Multiple Linear Regression (MLR) approaches as well as the Frequency premised segmentation algorithm is presented in the East Godavari district of Andhra Pradesh **[12]**.

Numerous studies have endeavored to handle forecasts using an Artificial Neural Network (ANN).Among the available neural networks, the Back Propagation Neural Network (BPNN) is a commonly used network design. In the backpropagation algorithm, there are three layers: intake, outcome, and camouflaged. The parameters were amended based on the amount of inaccuracy **[13]**. Multiple Linear Regression (MLR) and neural networks were used as forecasting analytics for potato yield based on data from numerous tests performed over three years **[14]**.

The intricate relationships and significant non - linearity between crop yields and multiple predictor parameters are solved using artificial neural networks (ANN). They are simple to implement and contain objective mathematical operations rather than interpretive guidelines.Additionally, they give exact outputs for circumstances that are not mentioned in the input.They also don't necessitate any pre-existing affiliation and could be generated used in public datasets.A most beneficial strategy to extracting information through imprecise, ambiguous, and semi-material is regarded to be ANN.In a variety of scenarios, such as yield output predictions, ANN has proven to be a reliable platform. **[15]** An ANN-based corn production forecasting scheme is presented, with soil and meteorological conditions as input elements.To fine-tune the product, convolution neural networks (CNN) and deep neural networks (DNN) or classification models can be used.

Support Vector Machine (SVM) established by **[16]** can be used to make predictions.In a high- or infinite-dimensional space, SVM generates a higher dimensional space or groupof higher dimensional space that could be used for extrapolation, segmentation, and other applications.In nonlinear circumstances, SVM plots the information into a greater feature set, where quadratic algorithms can be implemented, using a kernel methodology.SVM has a significant advantage more than other statistical techniques such as ANN in terms of ease of use because it demands only a few components to calibrate the model.

[17] Proposes implementing SVM to classify soils to suggest acceptable crops for the land.In SVM modeling, identifying the optimization techniques, hyper components, and penalty coefficient is critical. **[18]** Suggests using SVM to assess aerial hyperspectral images generated over a maize field to forecast crop output, biomass, plant height, and leaf greenness. When evaluated to a stepwise regression procedure, the findings were ascertained to be superior.

Measurements

It's indeed apparent from the preceding section that several machine learning techniques are implemented to forecast crop production.The numerous computational intelligence algorithms or accessible to usage, the complexity in the implementation, and the reliability of the forecasting model determine which algorithm is used.Different statistics, such as Mean Absolute Error and Root Mean Squared Error, are used to confirm the effectiveness of classifiers.

The Average squared error of estimations is measured by the Mean Squared Error (MSE).That is, this quantifies the square of the difference between the expected and leads to better before aggregating them. The MSE computation is shown in Equation 1.

$$\boldsymbol{Mean\ Squared\ Error} = \frac{\mathbf{1}}{\boldsymbol{M}}\sum_{j=1}^{M}(\boldsymbol{b}-\overline{\boldsymbol{b}})^{2} \qquad (\mathbf{1})$$

Mean Square Error with a limited quality indicates that the predictor is accurate. The MSE value for a perfect discriminator would be 0.

Root Mean Squared Error (RMSE) is the square root of MSE. The Root Mean Square Error (RMSE) is computed as follows in Equation 2.

$$\boldsymbol{Root\ Mean} = \sqrt{\frac{\mathbf{1}}{\boldsymbol{M}}\sum_{j=1}^{M}(\boldsymbol{b}-\overline{\boldsymbol{b}})^{2}} \qquad (\mathbf{2})$$

The aggregate of comparative discrepancies between the target and projected values is computed as the Mean Absolute Error (MAE). This is written as in Equation 3.

$$\boldsymbol{Mean\ Absolute\ Error} = \frac{\mathbf{1}}{\boldsymbol{M}}\sum_{j=1}^{M}|\boldsymbol{b}-\overline{\boldsymbol{b}}|^{2} \qquad (\mathbf{3})$$

The purpose of the present study would have been to create websites to examine the effect of meteorological variables and crop yield in Madhya Pradesh's selected districts.The districts were decided based on the quantity of land cultivated in that particular crop.The top five districts with the largest determining crop area were selected depending on these criteria.Products for the study were identified based on the most frequent crops in the selected district.Soybean, corn, paddy, and barley are just a few of the products that have been selected. Information sources were used to determine the production of these products over 20 years. Additionally, for the preceding year's meteorological parameters, such as rainfall, median temperatures, and potential evapotranspiration, reference materials were used to determine the prevalence of rainy days and cloud cover.The optimization techniques comprised treating numbers that exceeded the requirements into one kid as well as the rest as the other. Also, it deals with vanishing the primary key. In pseudo-code, the following is the generic algorithm of creating tree structure:

Let R be a collection of training images, each with a known class label.Every sample is essentially a tuple. The category of training samples is determined by one attribute. Let's pretend there are m different types of categories. Let Rinclude R_j examples of class S_j for J =1.., n, With probability rj/r,class $R_j$includes any sample, while s is the entire number of samples in set R. The following is a list of the necessary information to categorize a specimen using Equation 4.

$$\boldsymbol{J(R_1, R_2, \dots \dots \dots R_m) = -\sum_{j=1}^{n} \frac{R_j}{R} \log \frac{R_j}{R}} \quad \textbf{(4)}$$

Using a property R with values $R_1, R_2, \dots \dots \dots R_m$, S can be partitioned into subsets a_1, a_2,..., a_v, and R_k comprises the specimens in R which have value aj of A.Allow R_k to keep $r_{jk}.C_i$ specimens in his possession. It’s expected to improve the availability on A's segmentation is the entropy of R. The following is the absolute value shown in Equations 5 and 6.

$$\boldsymbol{F(Attribute) = \sum_{k=1}^{u} \frac{R_{1j} + \dots R_{nj}}{R} J(R_{1j}, \dots R_{nj})} \quad \textbf{(5)}$$

$$\boldsymbol{Gain(attribute) = J((R_1, R_2, \dots \dots \dots R_m) - F(Attribute)} \quad \textbf{(6)}$$

They can estimate the training learning in each of the different reliability specimens to S using this technique for significance evaluation.The most distinguishing characteristic in the given collection is the attribute and the maximum information acquisition.As a consequence of estimating the information acquisition for each quality, we can rank the qualities. This rating can be used to perform a significance investigation to choose which qualities to utilize in a concept explanation.

APPROACH AND DISCUSSIONS

According to the literature, the forecasting model takes into account whether contributions such as rainfall, temperature, and tenderness, as well as non-weather contributions such as soil moisture, pH, salts in soil (calcium, magnesium nitrogen, phosphate, potassium, organic carbon, sulfur, and so on), type of crop and grain category.The original data were first normalized to eliminate any erroneous or improperly formatted information.The information generated is then sent into the estimation method.Estimation techniques based on Support Vector Machines and Artificial Neural Networks have been considered satisfactory for agricultural production, estimation when evaluation metrics of multiple machine learning models are examined. Additionally, according to the literature, a considerable amount of data is constantly produced as a result of digital improvements in the field of farming.As a result, farm information has penetrated the large datasets world **[19]**. With this in mind, a technique for assessing crop yield forecasting in a large data set is proposed in this research. **Figure 1** depicts the proposed methodology as a block diagram.

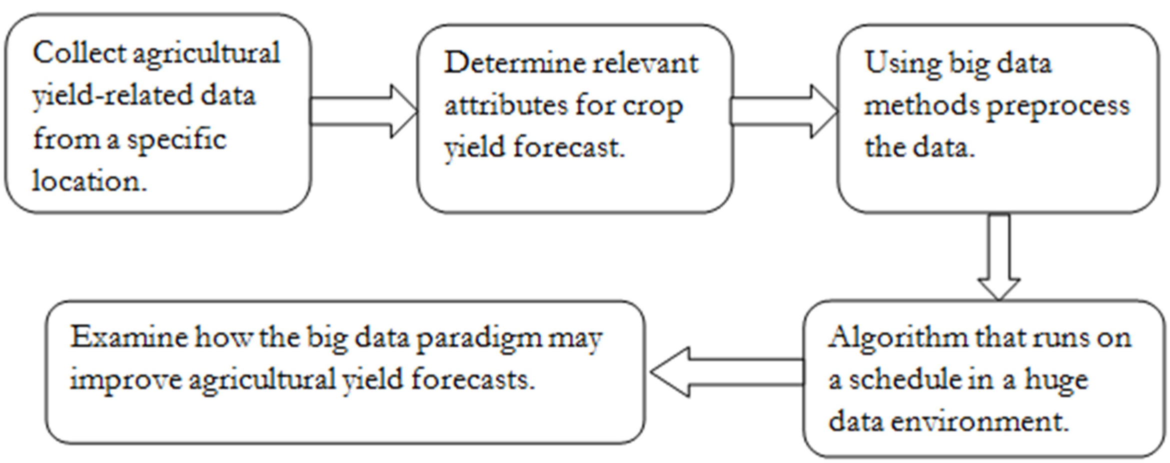

Figure 1: Using large data tools and techniques to estimate the yield of black gram

Figure 1 as shown in big data approaches were developed to be used at two stages: I preparation and (ii) forecasting.It is important to improve forecasting accuracy by spreading computation across numerous data nodes using a parallel programming methodology.As a result, it is conducted to determine how massive data capabilities might be used to optimize crop yield forecasts.

Only meteorological parameters were used to forecast agricultural output in this study, even though crop productivity is determined by numerous other input parameters such as irrigation, crop management, pesticide treatment, and so on.This is attributable to a shortage of this information on a district-by-district basis, which led to the development of a model that can roughly anticipate crop yields using meteorological characteristics, as this would areas of competence determine on grain buffer stocks, predefined threshold prices, and so on.As a result, the built program's decision rules were utilized to authenticate this programmer by forecasting the quantities of various productions in all of the specified districts using the measured values.The reliability of the forecast was confirmed by observing the expected and observed yields. The developed programmer has been confirmed for each crop; nonetheless, the criteria and confirmation reliability of the results for one crop (soybean) are reported in this study.

The decision rules developed based on the model for soybean crop in Dewas district are:

i. If Insulations <43 days & rain 888.23mm then Harvest < 800 kg/ha.
ii. If Insulations >43 days & rain 888.23mm then Harvest > 800 kg/ha.
iii. If Insulations > 43 days & Minimum temperature < 18.62 degree Celsius then Yield < 800kg/ha.
iv. If Insulations > 43 days & Minimum temperature > 18.42 degree Celsius then Yield > 800kg/ha.

Table 1 shows the expected estimates of the district's greatest critical metrics based on the preceding evaluation criteria.

Out of 20 years of data, 18 years of estimations were correct and two years were incorrect, putting the developed model's prediction performance at 90% in the case of soybeans.Similar comparisons were conducted to all the crop production and districts, as well as the total reliability of the built model are shown in Table 2based on the results. A definite pattern from each product being affected by a certain climatic characteristic was discovered using web-based software developed to estimate crop production based on the responses of different climatic statistics. **Table 2** shows the average reliability attained for each crop in various districts, as well as the prediction performance of the constructed model for various crops.

For the specified crops and districts, the generated model's forecast reliability ranged from 76 to 90 percent.Based on these findings, the aggregate generalization ability of the created model is 82.00 percent.Because of its great demand forecasting, the created model can be used by regulators to decide policy far ahead of crop production.

Table 1: Adeveloped model's soybean crop prediction performance

Insulations/Days	Rain/mm	Minimum Temperature	Harvest, kg/ha	Scheduled Yield (kg/ha)	Accuracy
43.85	942.20	18.65	900.6	>800	Y
40.45	1280.65	18.35	921.5	>800	Y
42.65	1032.25	18.42	935.4	>800	Y
41.7	1121.62	19.05	925.8	>800	N
44.52	1076.76	18.72	958.4	>800	Y

Table 2: Reliability

List	The prediction accuracy on average
Barley	78%
Castor Bean	58%
Flax	67%
Mustard	82%

CONCLUSION

A proper investigation of crop forecasting is conducted. Various machine learning algorithms for crop prediction have been established in the literature.Classification techniques methods' quality metrics, including root, mean square error, are also monitored.It is conducted to determine the impact of information methodologies on forecasting, in addition to machine learning algorithms. For this, a conceptual framework is suggested. The proposed approach is currently being implemented.The generated homepage is user-friendly, and the accuracy rate is greater than 75% in all the crops and districts studied, demonstrating higher forecast accuracy. Any user can utilize the user-friendly web page built for forecasting crop production by entering climatic data for their preferred crop.

REFERENCES

[1] Kannan E, Sundaram S. Analysis of trends in India's Agricultural Growth. Bangalore, India; 2011.

[2] Chand R, Kumar P, Kumar S. Total Factor Productivity and Returns to Public Investment on Agricultural Research in India §. Agricultural

Economics Research Review. 2012; 25(2): 181-94.

[3] Tanveer A, Ikram RM, Ali HH. Crop Rotation: Principles and Practices. InAgronomic Crops 2019 (pp. 1-12). Springer, Singapore.

[4] Dr.P.Sivakumar, "Analytical framework to build predictive and optimization function from manufacturing industry sensor data using cross-sectional sharing", Big Data,2021 (SCI)

[5] Dr.P.Sivakumar, "Improved Resource management and utilization based on a fog-cloud computing system with IoT incorporated with Classifier systems", Microprocessors and Microsystems, Jan 2021 (SCI).

[6] Ranjeeth, S., Latchoumi, T. P., & Paul, P. V. (2020). Role of gender on academic performance based on different parameters: Data from secondary school education. Data in brief, 29, 105257.

[7] Venkata Pavan, M., Karnan, B., &Latchoumi, T. P. (2021). PLA-Cu reinforced composite filament: Preparation and flexural property printed at different machining conditions. Advanced Composite Materials, https://doi. org/10.1080/09243046.2021, 1918608.

[8] Bakhshipour A, Jafari A. Evaluation of support vector machine and artificial neural networks in weed detection using shape features. Computers and Electronics in Agriculture. 2018 Feb 1; 145: 153-60.

[9] Park JG, Kim KJ. Design of a visual perception model with edge-adaptive Gabor filter and support vector machine for traffic sign detection. Expert Systems with Applications. 2013 Jul 1; 40(9): 3679-87.

[10] Blum MG, François O. Non-linear regression models for Approximate Bayesian Computation. Statistics and computing. 2010 Jan 1; 20(1): 63-73.

[11] Gasparrini A, Armstrong B, Kenward MG. Multivariate meta-analysis for non-linear and other multi-parameter associations. Statistics in medicine. 2012 Dec 20; 31(29): 3821-39.

[12] Billings SA, Chen S. Extended model set, global data and threshold model identification of severely non-linear systems. International Journal of Control. 1989 Nov 1; 50(5): 1897-923.

[13] Erb RJ. Introduction to backpropagation neural network computation. Pharmaceutical research. 1993 Feb; 10(2): 165-70.

[14] Mujahidin I, Prasetya DA, Nachrowie SA, Arinda PS. Performance tuning of spade card antenna using the mean average loss of backpropagation neural network. Int. J. Adv. Comput. Sci. Appl. 2020.

[15] Shaik NB, Pedapati SR, Taqvi SA, Othman AR, Dzubir FA. A feed-forward backpropagation neural network approach to predict the life condition of a crude oil pipeline. Processes. 2020 Jun; 8(6): 661.

[16] Wang L, Wang P, Liang S, Zhu Y, Khan J, Fang S. Monitoring maize growth on the North China Plain using a hybrid genetic algorithm-based back-propagation neural network model. Computers and Electronics in Agriculture. 2020 Mar 1; 170: 105238.

[17] Li X, Cheng X, Wu W, Wang Q, Tong Z, Zhang X, Deng D, Li Y. Forecasting of bioaerosol concentration by a Back Propagation neural network model. Science of The Total Environment. 2020 Jan 1; 698: 134315.

[18] Lee H, Wang J, Leblon B. Using linear regression, Random Forests, and Support Vector Machine with unmanned aerial vehicle multispectral images to predict canopy nitrogen weight in corn. Remote Sensing. 2020 Jan; 12(13): 2071.

[19] Misra NN, Dixit Y, Al-Mallahi A, Bhullar MS, Upadhyay R, Martynenko A. IoT, big data and artificial intelligence in agriculture and food industry. IEEE Internet of Things Journal. 2020 May 29.