International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

IJIEMR Transactions, online available on 12th September 2024. Link

https://ijiemr.org/downloads.php?vol=Volume-13&issue=issue09

DOI: http://doi.org/10.36893/IJIEMR/2024/V13/I09/16

Title: " CROP PREDICTION BASED ON CHARACTERISTICS OF THE AGRICULTURAL ENVIRONMENT USING VARIOUS FEATURE SELECTION TECHNIQUES AND CLASSIFIERS "

Volume 13, ISSUE 09, Pages: 153 - 160

Paper Authors

Mr.Karamsetty Shouryadhar,S. Sai Namitha,R. Lohitha,T. Rani

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

CROP PREDICTION BASED ON CHARACTERISTICS OF THE AGRICULTURAL ENVIRONMENT USING VARIOUS FEATURE SELECTION TECHNIQUES AND CLASSIFIERS

[1]Mr.Karamsetty Shouryadhar, [2]S. Sai Namitha,[3]R. Lohitha,[4]T. Rani
[1]Assistant Professor,[234]Students
Department Of CSE
Malla Reddy Engineering College for Women

ABSTRACT
Research in agriculture is expanding. Specifically, crop prediction is important in agriculture and mostly depends on soil and environmental factors including temperature, humidity, and rainfall. Previously, farmers could choose the crop to be grown, track its development, and select when it was ready for harvest. However, the agricultural community finds it difficult to continue doing so in the modern day due to the quick changes in environmental circumstances. As a result, machine learning approaches have dominated prediction tasks in recent years. This study has employed many of these techniques to estimate agricultural production. In order to guarantee that a particular machine learning (ML) model operates with a high degree of accuracy, effective feature selection techniques must be used to transform the raw data into a dataset that is machine learning-friendly and readily computed. Only data characteristics that are very relevant in predicting the model's final output should be included, since this will cut down on redundancies and improve the accuracy of the ML model. In order to guarantee that only the most relevant characteristics are included in the model, the concept of optimum feature selection is created. Compiling all features from the raw data and then lumping them together without considering their significance to the model-building process would make our model too complex. Moreover, adding characteristics that don't add much to the ML model will make it more complicated in terms of time and space, which would reduce the output accuracy of the model. The findings show that an ensemble method outperforms the current classification strategy in terms of prediction accuracy.

1. INTRODUCTION

In agriculture, crop prediction is a complex process [1], and many models have been put out and examined in this regard. Given both biotic and abiotic variables influence crop production, the challenge necessitates the utilization of diverse datasets [2]. The components of the environment known as biotic factors are those that arise as a direct or indirect consequence of living things (plants, animals, insects, predators, and microbes) influencing other living things. Anthropogenic variables (fertilization, plant protection, irrigation, air, water, and soil pollution, etc.) are also included in this category. These variables may result in internal flaws, form flaws, and modifications in the chemical makeup of the plant yield, among many other variations in crop yield. Both biotic and abiotic variables affect plant development and quality as well

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

as how the environment is shaped.There are three categories of biotic factors: chemical, physical, and other. The following physical factors have been identified: soil type, topography, soil rockiness, atmosphere, water chemistry, particularly salinity; mechanical vibrations (noise, vibration); radiation (e.g., ionizing, electromagnetic, ultraviolet, and infrared); climatic conditions (temperature, humidity, air movements, and sunlight). The chemical components include nitrogen oxides and derivatives, fluorine and its compounds, lead and its compounds, cadmium and its compounds, nitrogen fertilizers, pesticides, and carbon monoxide. Priority environmental toxins include sulfur dioxide and its derivatives and PAHs. The others include asbestos, dioxins and furans, arsenic, mercury, and toxins [3]. In addition, bedrock, relief, temperature, and water conditions are biotic elements that impact the property. Factors that contribute to soil formation and agricultural value have a variety of effects [4].

Crop production prediction is neither straightforward nor easy. According to Myers et al. [5] and Muriithi [6], the approach for forecasting the area under cultivation is a collection of statistical and mathematical tools helpful in an iterative and improved optimization process. It is also useful for designing, developing, and formulating new goods as well as enhancing those that already exist. Numerical data must be in possession for statistical analysis to be performed or presented. On the basis of them, inferences about a variety of events are established, and legally enforceable economic judgments may then be taken. Muriithi [6] asserts that the more accurately you can characterize a phenomenon in numerical terms, the more you can say about it. Moreover, as data accuracy increases, you may also receive more precise information and make more correct judgments.

The evaluation of agroclimatic parameters in relation to influencing the production of winter plant species, mostly grains, is the largest challenge in the temperate temperature zone. The primary determinant of wintering yield is the quantity and frequency of days in the wintering period with temperatures above 0°C and 5°C, as well as access to days with a temperature above 5°C. Many of these may be calculated using publicly available data, producing regression statistics expressed in years. created models to evaluate the circumstances and determine if state policy in the area of grain market intervention should be put on trial. predictions of agrometeorological parameters is necessary for effective productivity predictions. Specific issues might arise because of these components' variability [7]. This problem has been tackled by several researchers, with differing degrees of success. [8]_[10].

Grabowska et al. [9] used meteorological models and three climate change scenarios for Central Europe (HadCM3, GFDL, and E-GISS model) to anticipate narrow-leaf lupine yields for 2050–2060. The determination coefficient R2, corrected coefficient of determination R2adj, standard error of estimation, and the coefficient of determination R2pred—all of which were computed using the Cross Validation procedure—were used to evaluate the _t of the models. The chosen

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

formula was used to predict the yield of lupines when the amount of CO2 in the atmosphere was doubled. According to these scientists, depending on the station's location, different weather parameters had different effects on the yield of narrow-leaved lupine. Most of the time, rainfall during the _owering - technical maturity period and temperature (maximum, average, and lowest) at the start of the growing season had a major impact on the yield. Research has shown that the anticipated changes in climate would positively impact the output of lupines. The most favorable scenario was HadCM3, and the simulated protability was greater than the observed protability from 1990 to 2008.

In order to anticipate crop yields in Poland, Dšbrowska-Zieli«ska et al. [8] evaluated the value of plant biophysical characteristics that were computed from the ranges of rejected electromagnetic radiation captured by the new generation satellites Sentinel-2 and Proba-V. Ground measurements were done in 2016–2018 in arable areas under the GEO Joint Experiment of Crop Assessment and Monitoring (JECAM) global crop monitoring network. Sentinel-1 and RadarSat-2 optical and radar pictures were used to classify crops. The standard Biomass and Evapo transpiration PRO model was used to anticipate the biomass size of winter wheat cultivation by simulating its development. Received a high accuracy of 94% of the predicted biomass size using actual biomass.

2. LITERATURE SURVEY

Predicting Rice Yield using Support Vector Machines, 2019. These days, support vector machines, or SVMs, are often used in the computer software industry. They are often used within the provision as they are able to procreate. The goal of this study is to build category fashions for Indian rice yield prediction that are mostly based on SVM. Experiments have been conducted using the kernel polynomial function of SVM education, okay-cross validation, and the multiple classification approach. For this research, statistics on rice production in India were given by the Department of Economics and Statistics of the Indian Department of Agriculture. The four-year relative average growth was predicted with a high accuracy of 75.06% using the four-triple cross-validation approach. The tests conducted in this article make use of MATLAB software.

Agricultural coverage has a major role in both economic development and the preservation of American agriculture's food supply. The choice of which crop or crops to grow is one of the major issues in agricultural planning. It is influenced by many variables, including market price, government coverage, and production rate. Several academics have studied crop forecasting, soil classification, crop class, and climate forecasting in order to apply statistical methodologies or system learning approaches to agricultural planning. When there are many possibilities for when to plant and utilize available land, choosing a crop becomes difficult. This research uses a strategy called the Crop Selection Method (CSM) to solve the crop selection issue, optimize crop internet at a certain moment in the yield duration, and eventually gain the highest financial increase of the site.

Farm Data System for Machine Learning, 2020. India is the second most renowned nation in the United States, as far as we know. Most people in the globe, including those in India, are employed in agriculture. Farmers use fertilizers in unknown quantities without sufficient knowledge of the kind and quantity of fertilizer, and they constantly plant the same crops without trying with alternative crop types. Consequently, it degrades the top layer, increases soil acidity, and has an impact on production all at once. Consequently, we created a system that boosts farmers' production levels via the use of device learning algorithms. Our approach will recommend the best crop for a given soil based just on the composition of the soil and climatic parameters. Additionally, the device offers details on the kind and amount of nutrients required for seed growth. Thus, by using our technology, farmers may grow new plant kinds, boost yields, and stop soil contamination.

Provide a summary of how machine learning techniques are being used in 2019. Agriculture is one of the most significant, yet least lucrative, occupations in India. Machine learning has the potential to improve profitability and encourage more agricultural production by growing the most desired crops. Through the use of various system analysis approaches, the aim of this work is to produce a yield forecast. The overall effectiveness of the methods is compared using the mean absolute errors. Machine learning algorithms will assist farmers choose which crop to cultivate in order to optimize output by taking into account factors like area, temperature, rainfall, and more.

3. EXISTING SYSTEM

You and colleagues [15] proposed a flexible and accurate method to predict yields using publicly available remote sensing data.

There are three areas in which the technique improves upon current practices. First, a functional approach is proposed using a distant detecting network. Next, a new dimensionality reduction process is introduced that combines long-term memory with a convolutional neural network (CNN). Lastly, the spatio-transient structure of the data is examined and improved in accuracy using a Gaussian process. Anantha et al. [16] used an association ensemble model with majority voting to create a recommendation system. When determining the best suitable crop, taking into account soil factors, the random tree, Chi-square Automatic Interaction Detection (CHAID), kNN, and Naive Bayes (NB) are used as learners. The findings demonstrate good accuracy and potency. The categorized picture produced by these methods is made up of mathematical information applied to ground truth. Moreover, it includes information on agricultural output by state and district as well as the square measure's characteristics for weather and crop yield.

All of the aforementioned methods are used to forecast certain crop yields under particular conditions. A forecasting model

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

for agricultural yield production was created by Rale et al. [17] using RF regression and the default parameters.

Fernando et al. [19] examined data on the yearly output of coconuts in a specific location from 1971 to 2001 and evaluated the economic implications. According to the report, the economy lost almost US $50 million as a result of the crop scarcity. An estimating method for rice yield prediction was developed by Ji et al. [20]. The goal of the research was to ascertain how well artificial neural networks (ANN) could forecast rice output in hilly areas. It evaluated the ANN's effectiveness in relation to biological parametric fluctuations and contrasted its performance with several bilinear regression models. Following the guidelines established by the National Agricultural Statistics Service (NASS) and the Cropland Data Layer (CDL), Boryan et al. [21] proposed a decision tree-based technique to represent publicly accessible state-level crop cover groups using ground truth gathered during the June Agricultural Survey. The NSS CDL program is described in the planned work.

It provides details on handling tactics, order and approval processes, accuracy assessment, CDL item specifics, and the process for estimating product costs. Landsat was suggested by Hansen and Loveland [22] as a way to get satellite pictures for environmental remote sensing.

Disadvantages

- ❖ The RECURSIVE FEATURE ELIMINATION (RFE) feature of the system is not implemented.
- ❖ The system is not put into use strategies for sampling used in preprocessing to optimize prediction performance and balance the dataset.

4. PROPOSED SYSTEM

Boruta is a classification technique based on random forests [38] that uses decision trees to vote on flexible, impartial, and indistinct classifiers. By estimating the loss of classification accuracy brought on by the random permutation of characteristics inside objects, the significance of a feature is determined. To quantify average changes in mean accuracy loss across crops, the average loss of accuracy is divided by the standard deviation to generate the Z score. This is done by calculating the average and standard deviation of the accuracy loss.

For every tree, a `shadow' property is created by reordering the original attribute values among items at random. By examining every characteristic in the system, the significance of each attribute is ascertained. Since the oscillations are unpredictable, the most significant qualities are shown by referring to the shadow attributes. The degree of precision is obviously highly dependent on the shadow properties. To get the best outcomes, the values will be continually reshuffled. The stages of the Boruta algorithm are as follows: 1. Five shadow attributes are always added to the data system, which is expanded by attaching duplicates of every shadow attribute.
2. To eliminate any association with the answer, the additional characteristics are mixed in with the original attribute.

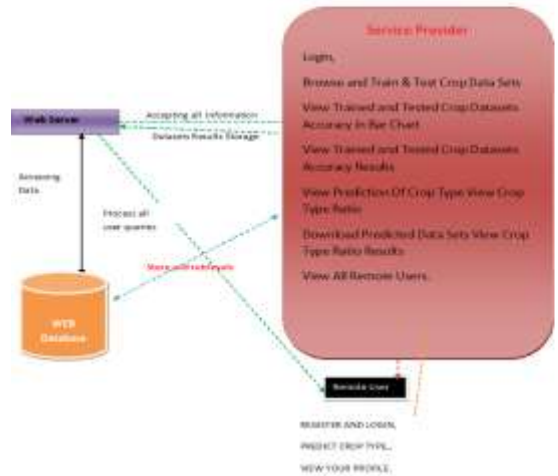
3. Using the widely used information system, a random forest technique is used to get the Z score.
4. After calculating the Maximum Z Score Attributes (MZSA), a "hit" is given to any attribute whose value exceeds the MZSA.
5. The MZSA is used to conduct a two-sided test of equality for qualities whose significance is unclear.
6. Features whose relevance is much less than the MZSA are labeled as "unimportant" and removed from the database forever.
7. "Important" attributes are those whose significance is noticeably greater than the MZSA.
8. The information system is therefore free of shadow properties.
9. Until each characteristic is assigned a priority, the procedure is repeated.

Advantages

- ➢ The RFE method begins with the whole dataset and is a wrapper feature selection strategy. Using a ranking mechanism that is essential to the RFE methodology, the dataset is ranked from best to worst depending on which salient characteristics are chosen.
- ➢ The primary benefit of the RFE over other techniques is that it removes features exclusively based on their performance and objectively confirms each feature's function in processing the model's output.

5. SYSTEM ARCHITECTURE

Architecture Diagram



6. MODULES

Service Provider

The Service Provider must provide a valid user name and password to log in to this module. Following a successful login, one may do a number of tasks, including Examine and Test and Train Crop Data Sets, View the accuracy of trained and tested crop datasets in a bar chart; view the accuracy results of trained and tested crop datasets; view the crop type prediction; view the crop type ratio; download the predicted data sets; view the results of the crop type ratio; and view all remote users.

View and Authorize Users

The administrator may see a list of all enrolled users in this module. The administrator may see user information here, including name, email address, and address, and they can also approve people.

Remote User

There are n numbers of users present in this module. Prior to beginning any actions, the user must register. The user's information is saved in the database when they register. Upon successful registration, he must use his permitted user name and password to log in. Upon successful login, the user may do several tasks such as registering and logging in, predicting the crop type, and seeing their profile.

CONCLUSION

Crop forecasting is a challenging endeavor in agriculture. This study predicted the magnitude of plant cultivation yields using a variety of feature selection and classification methods. The findings show that an ensemble approach outperforms the current classification method in terms of prediction accuracy. On a farm and national level, the structure of the planting of energy crops, such as potatoes and grains, may be planned by forecasting the area of these crops. Utilizing contemporary forecasting methods might provide quantifiable financial advantages.

REFERENCES

[1] R. Jahan, ``Applying naive Bayes classi_cation technique for classi_cation of improved agricultural land soils,'' *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 6, no. 5, pp. 189_193, May 2018.

[2] B. B. Sawicka and B. Krochmal-Marczak, ``Biotic components in_uencing the yield and quality of potato tubers,'' *Herbalism*, vol. 1, no. 3, pp. 125_136, 2017.

[3] B. Sawicka, A. H. Noaema, and A. Gáowacka, ``The predicting the size of the potato acreage as a raw material for bioethanol production,'' in *Alternative Energy Sources*, B. Zdunek, M. Olszáwka, Eds. Lublin, Poland: Wydawnictwo Naukowe TYGIEL, 2016, pp. 158_172.

[4] B. Sawicka, A. H. Noaema, T. S. Hameed, and B. Krochmal-Marczak, ``Biotic and abiotic factors in_uencing on the environment and growth of plants,'' (in Polish), in *Proc. Bioró»norodno±¢ .rodowiska Znaczenie, Problemy, Wyzwania. Materiaªy Konferencyjne*, Puªawy, May 2017. [Online]. Available: https://bookcrossing.pl/ksiazka/321192

[5] R. H. Myers, D. C. Montgomery, G. G. Vining, C. M. Borror, and S. M. Kowalski, ``Response surface methodology: A retrospective and literature survey,'' *J. Qual. Technol.*, vol. 36, no. 1, pp. 53_77, Jan. 2004.

[6] D. K. Muriithi, ``Application of response surface methodology for optimization of potato tuber yield,'' *Amer. J. Theor. Appl. Statist.*, vol. 4, no. 4, pp. 300_304, 2015, doi: 10.11648/j.ajtas.20150404.20.

[7] M. Marenych, O. Verevska, A. Kalinichenko, and M. Dacko, ``Assessment of the impact of weather conditions on the yield of winter wheat in Ukraine in terms of regional,'' *Assoc. Agricult. Agribusiness Econ. Ann. Sci.*, vol. 16, no. 2, pp. 183_188, 2014.

[8] J. R. Ol¦dzki, ``The report on the state of remotesensing in Poland in 2011_2014,'' (in Polish), *Remote Sens. Environ.*, vol. 53, no. 2, pp. 113_174, 2015.

[9] K. Grabowska, A. Dymerska, K. Poáarska, and J. Grabowski, ``Predicting of blue lupine yields based on the selected climate change scenarios,'' *Acta*

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Agroph., vol. 23, no. 3, pp. 363_380, 2016.
[10] D. Li, Y. Miao, S. K. Gupta, C. J. Rosen, F. Yuan, C. Wang, L. Wang, and Y. Huang, ``Improving potato yield prediction by combining cultivar information and UAV remote sensing data using machine learning," *Remote Sens.*, vol. 13, no. 16, p. 3322, Aug. 2021, doi: 10.3390/rs13163322.
[11] N. Chanamarn, K. Tamee, and P. Sittidech, ``Stacking technique for academic achievement prediction," in *Proc. Int. Workshop Smart Info-Media Syst.*, 2016, pp. 14_17.
[12] W. Paja, K. Pancerz, and P. Grochowalski, ``Generational feature elimination and some other ranking feature selection methods," in *Advances in Feature Selection for Data and Pattern Recognition*, vol. 138. Cham, Switzerland: Springer, 2018, pp. 97_112.
[13] D. C. Duro, S. E. Franklin, and M. G. Dubé, ``A comparison of pixelbased and object-based image analysis with selected machine learning algorithms for the classi_cation of agricultural landscapes using SPOT-5 HRG imagery," *Remote Sens. Environ.*, vol. 118, pp. 259_272, Mar. 2012.
[14] S. K. Honawad, S. S. Chinchali, K. Pawar, and P. Deshpande, ``Soil classi_cation and suitable crop prediction," in *Proc. Nat. Conf. Comput. Biol., Commun., Data Anal.* 2017, pp. 25_29.
[15] J. You, X. Li, M. Low, D. Lobell, and S. Ermon, ``Deep Gaussian process for crop yield prediction based on remote sensing data," in *Proc. AAAI Conf. Artif. Intell.*, 2017, vol. 31, no. 1, pp. 4559_4565.