JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Crop Prediction Based on Characteristics of the Agricultural Environment Using Various Feature Selection Techniques and Classifiers

Mr. RAMBABU ATMAKURI - Head, Department of CSE, Anurag College of Engineering (Aushapur, Ghatkesar, Telangana 501301)

Mr. RALLABANDI BHARGAVA - Student, Department of CSE, Anurag College of Engineering (Aushapur, Ghatkesar, Telangana 501301)

Ms. KODIGANTI HARIKA - Student, Department of CSE, Anurag College of Engineering (Aushapur, Ghatkesar, Telangana 501301)

Ms. GUMPARTHI DEVI GAYATHRI - Student, Department of CSE, Anurag College of Engineering (Aushapur, Ghatkesar, Telangana 501301)

Abstract

Agriculture is a growing field of research. In particular, crop prediction in agriculture is critical and is chiefly contingent upon soil and environment conditions, including rainfall, humidity, and temperature. In the past, farmers were able to decide on the crop to be cultivated, monitor its growth, and determine when it could be harvested. Today, however, rapid changes in environmental conditions have made it difficult for the farming community to continue to do so. Consequently, in recent years, machine learning techniques have taken over the task of prediction, and this work has used several of these to determine crop yield. To ensure that a given machine learning (ML) model works at a high level of precision, it is imperative to employ efficient feature selection methods to preprocess the raw data into an easily computable Machine Learning friendly dataset. To reduce redundancies and make the ML model more accurate, only data features that have a significant degree of relevance in determining the final output of the model must be employed. Thus, optimal feature selection arises to ensure that only the most relevant features are accepted as a part of the model. Conglomerating every single feature from raw data without checking for their role in the process of making the model will unnecessarily complicate our model. Furthermore, additional features which contribute little to the ML model will increase its time and space complexity and affect the accuracy of the model's output. The results depict that an ensemble technique offers better prediction accuracy than the existing classification technique.

1. INTRODUCTION

Crop prediction in agriculture is a complicated process [1] and multiple models have been proposed and tested to this end. The problem calls for the use of assorted datasets, given that crop cultivation depends on biotic and abiotic factors [2]. Biotic factors include those elements of the environment that occur as a result of the impact of living organisms (microorganisms, plants, animals, parasites, predators, pests), directly or indirectly, on other living organisms. This group also includes anthropogenic factors (fertilization, plant protection, irrigation, air pollution, water pollution and soils, etc.). These factors may contribute to the occurrence of many changes in the yield of crops, cause internal defects, shape defects and changes in the chemical composition of the plant yield. The shaping of the environment as well as the growth and quality of plants is influenced by abiotic and biotic factors. Abiotic factors can be divided into physical, chemical, and other. The recognized physical factors include: mechanical vibrations (vibration, noise), radiation (e.g., ionizing, electromagnetic, ultraviolet, infrared); climatic conditions (atmospheric pressure, temperature, humidity, air movements, sunlight); soil type, topography, soil rockiness, atmosphere, and water chemistry, especially salinity. The chemical factors

include: priority environmental poisons, such as sulfur dioxide and derivatives, PAHs; nitrogen oxides and derivatives, fluorine, and its compounds, lead and its compounds, cadmium and its compounds, nitrogen fertilizers, pesticides, carbon monoxide. The others are: mercury, arsenic, dioxins and furans, asbestos, and aflatoxins [3]. Abiotic factors also include bedrock, relief, climate, and water conditions - all of which affect its properties. Soil-forming factors have a diversified effect on the formation of soils and their agricultural value [4]. Predicting crop yields is neither simple nor easy. The methodology for predicting the area under cultivation is, according to Myers et al. [5] and Muriithi [6], a set of statistical and mathematical techniques useful in an evolving and improving optimization process. It also has important uses in design, development, and formulation new as well as improving existing products. Presentation or performance of statistical analysis requires the possession of numerical data. Based on them, conclusions are drawn as to various phenomena and further, on this basis, binding economic decisions can be made. According to Muriithi [6], the better you describe certain phenomena in terms of numbers, the more you can say about them, and with increasing data accuracy you can also obtain more accurate information and make more accurate decisions. The biggest problem in the temperate climate zone is assessment of agroclimatic factors in terms of shaping the yield of winter plant species, mainly cereals. The key factor influencing wintering yield, which provides access to days with a temperature over of 5◦ C, their number and frequency, and the number of days in the wintering period with temperatures above 0◦C and 5◦C. A number of these can be estimated on the basis of public statistics and yield regression statistics in years. Developed models for checking the situation that assess whether they want to be a probation of state policy in the field of intervention in the cereal market. Efficient forecasting of productivity requires forecasting of a agrometeorological factors. Aspects related to the variability of these factors may pose a particular problem [7]. Many researchers have dealt with this issue with varying degrees of success [8] [10].

2. LITERATURE SURVEY

Rakowski et al. [9] predicted narrow-leaf lupine yields for 2050-2060 using weather models and three climate change scenarios for Central Europe: E-GISS model, HadCM3 and GFDL. The fit of the models was assessed by means of the determination coefficient R2, corrected coefficient of determination R2adj, standard error of estimation and the coefficient of determination R2pred calculated using the Cross Validation procedure. The selected equation was used to forecast lupine yield under the conditions of doubling the CO2 content in the atmosphere. These authors stated that the influence of meteorological factors on the yield of narrow-leaved lupine varied depending on the location of the station. The temperature (maximum, average, minimum) at the beginning of the growing season, as well as rainfall during the flowering - technical maturity period, most often had a significant influence on the yield. It has been shown that the predicted climate changes will have a positive effect on the lupine yield. The simulated profitability was higher than that observed in 1990-2008, and HadCM3 was the most favorable scenario.

Dombrowska-Zielińska et al. [8] assessed the usefulness of plant biophysical parameters, calculated from the ranges of reflected electromagnetic radiation recorded by the new generation satellites Sentinel-2 and Proba-V, for forecasting crop yields in Poland. In 2016-2018, ground measurements were carried out in arable fields in the area included in the global crop monitoring network GEO Joint Experiment of Crop Assessment and Monitoring JECAM. Classification of crops was performed using optical and radar images Sentinel-1 and RadarSat-2. The PROtotypical model of Biomass and Evapotranspiration PRO was used to simulate the growth of winter wheat cultivation, to forecast its biomass size. Got high accuracy of 94% of the size of biomass modeled with real biomass.

Li et al. [10] found that accurate, high-resolution yield maps are needed to identify spatial patterns of yield variability, to identify key factors influencing yield variability, and to provide detailed management information in precision farming. Varietal differences may significantly affect the forecasting of potato tuber yields with the use of remote sensing technologies. These authors argue that improving potato crop forecasting with remote sensing of unmanned aerial vehicles (UAVs) by incorporating varietal information into machine learning methods has the best chance at present.

There are different challenges in this research area. Currently, crop prediction [11] models generate actual results that are satisfactory, though they could perform better. This paper attempts to propose an enhanced crop prediction model that addresses these issues. The prediction process [12] depends on the two fundamental techniques of feature selection [FS] and classification. Prior to the application of FS techniques, sampling

techniques are applied to balance an imbalanced dataset.

3. OVERVIEW OF THE SYSTEM

3.1 Existing System

The existing systems for crop prediction in agriculture utilize a variety of techniques, including statistical and mathematical models, machine learning algorithms, and remote sensing technologies. These systems leverage diverse datasets that capture biotic and abiotic factors affecting crop cultivation, such as weather conditions, soil properties, and environmental pollutants. Researchers have developed models that forecast crop yields based on historical data and projections of climate change scenarios, aiming to provide valuable insights for agricultural decision-making.

3.1.1 Disadvantages of Existing System

i. Inefficient feature selection techniques may lead to the inclusion of irrelevant data features, diminishing prediction accuracy.
ii. Limited adaptability to changing environmental conditions may result in less timely and precise predictions.
iii. Lack of robustness in handling complex interactions between biotic and abiotic factors can hinder the accuracy of crop yield forecasts.
iv. Suboptimal feature selection methods may result in the inclusion of irrelevant data features, reducing prediction accuracy.
v. Traditional techniques may struggle to adapt to dynamic environmental conditions, leading to less timely and accurate predictions.

3.2 Proposed System

The proposed system for crop prediction introduces the Boruta algorithm, a random forest-based classification technique designed to enhance feature selection. By extending the dataset with shadow attributes and employing iterative Z score calculations, Boruta accurately identifies the most relevant attributes affecting crop yield prediction. Through a rigorous process of significance testing and attribute elimination, the algorithm efficiently distinguishes between important and unimportant features, improving the precision of crop yield models. By leveraging the Boruta algorithm, the proposed system aims to optimize agricultural decision-making by providing valuable insights into the biotic and abiotic factors influencing crop cultivation.

3.2.1 Advantages of Proposed System

i. Enhanced feature selection accuracy through the utilization of the Boruta algorithm.
ii. Iterative approach ensures thorough evaluation of attribute importance, improving decision-making in agricultural practices.
iii. Optimization of feature selection captures complex interactions between biotic and abiotic factors, providing valuable insights for sustainable agricultural management.
iv. Utilization of the Boruta algorithm enhances feature selection accuracy, ensuring the inclusion of only the most relevant attributes for prediction.
v. Iterative evaluation process improves the robustness of attribute importance assessment, leading to more informed decision-making in agricultural management.
vi. Optimization of feature selection enables the capture of intricate relationships between biotic and abiotic factors, providing comprehensive insights for sustainable farming practices.
vii. Increased efficiency in identifying significant attributes enhances the overall performance and reliability of crop yield prediction models.
viii. The proposed system offers a systematic and data-driven approach, contributing to the advancement of precision agriculture and informed decision-making in crop cultivation.

3.3 Proposed System Design

In this project work, there are two modules and each module has specific functions, they are:

1. Service Provider Module
2. User Module

3.3.1 Service provider Module

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Login, Browse and Train & Test Crop Data Sets, View Trained and Tested Crop Datasets Accuracy in Bar Chart, View Trained and Tested Crop Datasets Accuracy Results, View Prediction Of Crop Type, View Crop Type Ratio, Download Predicted Data Sets, View Crop Type Ratio Results, View All Remote Users.

3.3.2 Remote User Module

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like Register And Login, Predict Crop Type, View Your Profile.

3.4 Architecture

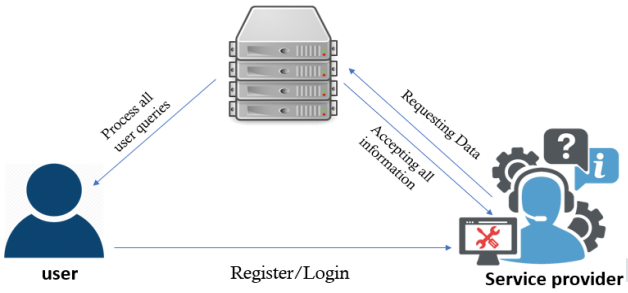
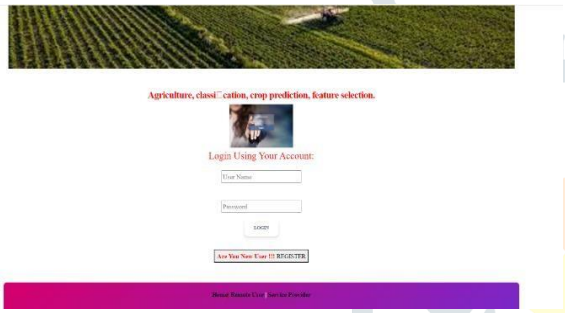
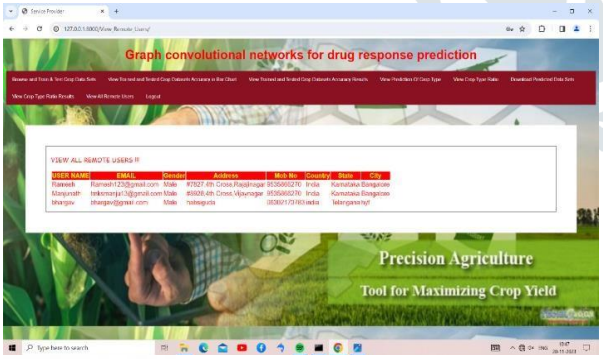
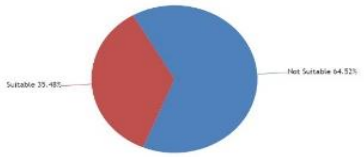


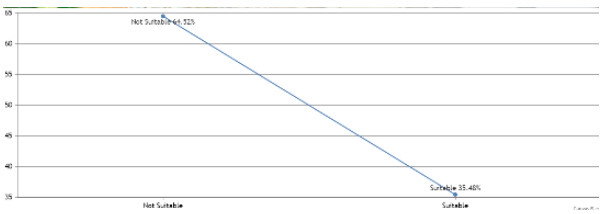
Fig 1: System Architecture

4. RESULT SCREEN SHOTS









Crop Prediction Found Ratio Details

Crop Prediction Type	Ratio
Not Suitable	64.51612903225805
Suitable	35.483870967741936

5. CONCLUSION

Predicting crops for cultivation in agriculture is a difficult task. This paper has used a range of feature selection and classification techniques to predict yield size of plant cultivations. The results depict that an ensemble technique offers better prediction accuracy than the existing classification technique. Forecasting the area of cereals, potatoes and other energy crops can be used to plan the structure of their sowing, both on the farm and country scale. The use of modern forecasting techniques can bring measurable financial benefits. To enhance crop prediction based on environmental characteristics, consider integrating advanced machine learning algorithms like Random Forest, Gradient Boosting, or Neural Networks. Additionally, explore feature selection techniques such as Recursive Feature Elimination or Principal Component Analysis to improve model efficiency. Stay updated on remote sensing technologies for accurate environmental data collection, and collaborate with domain experts to refine feature sets for better predictions. Regularly update your model with the latest agricultural and environmental research findings to ensure relevance and accuracy.

6. REFERENCES

[1] R. Jahan, ‘‘Applying naive Bayes classification technique for classification 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 influencing 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 influencing on the environment and growth of plants,’’

(in Polish), in Proc. Bioróżnorodność Środowiska Znaczenie, Problemy, Wyzwania. Materiały Konferencyjne, Puławy, 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. Olędzki, ‘‘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 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 classification 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 classification 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.

[16] D. A. Reddy, B. Dadore, and A. Watekar, ‘‘Crop recommendation system to maximize crop yield in ramtek region using machine learning,’’ Int. J. Sci. Res. Sci. Technol., vol. 6, no. 1, pp. 485–489, Feb. 2019.

[17] N. Rale, R. Solanki, D. Bein, J. Andro-Vasko, and W. Bein, ‘‘Prediction of Crop Cultivation,’’ in Proc. 19th Annu. Comput. Commun. Workshop Conf. (CCWC), Las Vegas, NV, USA, 2019, pp. 227–232.

[18] J. Jones, G. Hoogenboom, C. Porter, K. Boote, W. Batchelor, L. Hunt, P. Wilkens, U. Singh, A. Gijsman, and J. Ritchie, ‘‘The DSSAT cropping system model,’’ Eur. J. Agronomy, vol. 18, nos. 3–4, pp. 235–265, 2003.

[19] M. T. N. Fernando, L. Zubair, T. S. G. Peiris, C. S. Ranasinghe, and J. Ratnasiri, ‘‘Economic value of climate variability impact on coconut production in Sri Lanka,’’ in Proc. AIACC Working Papers, vol. 45, 2007, pp. 1–7.

[20] B. Ji, Y. Sun, S. Yang, and J. Wan, ‘‘Artificial neural networks for rice yield prediction in mountainous regions,’’ J. Agricult. Sci., vol. 145, no. 3, pp. 249–261, Jun. 2007.

[21] C. Boryan, Z. Yang, R. Mueller, and M. Craig, ‘‘Monitoring U.S. agriculture: The U.S. department of agriculture, national agricultural statistics service, cropland data layer program,’’ Geocarto Int., vol. 26, no. 5, pp. 341–358, 2011.

[22] M. C. Hansen and T. R. Loveland, ‘‘A review of large area monitoring of land cover change using Landsat data,’’ Remote Sens. Environ., vol. 122, pp. 66–74, Jul. 2012.

[23] D. K. Bolton and M. A. Friedl, ‘‘Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics,’’ Agricult. Forest Meteorol., vol. 173, pp. 74–84, May 2013.

[24] J. Dempewolf, B. Adusei, I. Becker-Reshef, M. Hansen, P. Potapov, A. Khan, and B. Barker, ‘‘Wheat yield forecasting for Punjab province from vegetation index time series and historic crop statistics,’’ Remote Sens., vol. 6, no. 10, pp. 9653–9675, Oct. 2014.

[25] H. D. Shannon and P. M. Raymond, ‘‘Managing weather and climate risk to agriculture in North America,’’ Central Amer. Caribbean, vol. 10, pp. 50–56, Dec. 2015.