© July 2024 | IJIRT | Volume 11 Issue 2 | ISSN: 2349-6002

Crop production optimisation: a blended technique using XGBoost and Simulated Annealing

Mallyala Nissie[1], Kasarla Sai Anirudh[2]
[1,2]*Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, India*

Abstract- **India's economy thrives mainly because of agriculture, which takes care of billions of people. Even so, it requires a new way to help improve crop production since there exist problems like over population, changing weather patterns or need for sustainable practices. The authors of this paper decided to introduce a different technique for boosting crop yields in different parts of India through a combination of Simulated Annealing (SA) and Extreme Gradient Boosting (XGBoost) models. The methodology aims at making crop management approaches better suited.**

Index Terms- **Farming, Machine learning, Stimulated Annealing, XGBoost**

1. INTRODUCTION

The underpinning of our general public, cultivating supports billions of individuals all around the world and is their essential kind of revenue. Antiquated individuals had the option to acclimate to their necessities since they collected food on their property. Thusly, different species, counting people, animals, and birds, develop and utilize normal crops. There is a developing requirement for effective fixes in the horticulture area because of elements including a populace blast, a sporadic environment, and the prerequisite for supportable methods. There are difficult issues confronting the world food framework now, and these issues ought to just deteriorate over the next forty years. Assuming there is sufficient will and subsidizing, progression might be made rapidly with current innovations furthermore, information. To track down imaginative responses, handling forthcoming issues will require more critical examination financing what's more, exceptional changes. Disintegration of biological systems, fatigue of regular assets, deteriorating ranch income, divided land possessions, and an absence of occupations in non-ranch ventures are a portion of the issues that exist today. Using new innovations is believed to be a urgent way to deal with bringing agrarian efficiency up later on. As opposed to overseeing whole fields in light of theoretical midpoints, accuracy cultivating perceives the differentiations between each site and changes the board procedures accordingly. The research being examined investigates an imaginative methodology to give ranchers reasonable information about which crops are ideal to develop given specific soil types and other basic attributes. Its will likely make a shrewd framework that ganders at significant factors and makes a dynamic plan that offers ranchers explicit guidance on what yields are best for various areas and seasons. This methodology uses current examination to convey modified and successful cultivating timetables to further develop decision-production for manageable and productive agrarian activities. The methodology starts with a cautious spotlight on information quality since solid and exact data fills in as the establishment for every logical interaction. An exhaustive period of information readiness and cleaning guarantees the trustworthiness of the bits of knowledge acquired by laying the basis for additional studies. An exploratory information examination (EDA) is led in view of the yield idea dataset to recommend crops that might be delivered in reasonable conditions while taking into account specific soil boundaries. The essential joining lies at the core of our examination. With the utilization of XGBoost, we can figure out which yields are generally appropriate for a specific region relying upon a few factors. This further develops crop choice accuracy, augmenting horticultural procedures for the particular locale and cultivating successful and supportable cultivating. A strong AI framework called XGBoost inspects many factors, counting soil type, environment, and past information, to convey exact suggestions for the best harvests to fill in a specific region. Using this front line prescient demonstrating, ranchers can pursue very much educated choices on practical and useful horticultural strategies by acquiring a intensive handle of yield similarity. This imaginative methodology advances more prominent creation and perseverance despite changing hardships in the farming business, addressing a significant stage towards accuracy cultivating. To decide the ideal time for developing explicit harvests, the utilization of reproduced toughening arises as a significant solution. Mimicked tempering, a metaheuristic calculation propelled by tempering cycles in metallurgy, adds to the streamlining of establishing plans. By methodicallly investigating different worldly potential outcomes and adjusting to evolving conditions, recreated tempering guides in recognizing the most positive time frames for crop cultivation.This approach upgrades accuracy in rural preparation, permitting ranchers to synchronize planting with natural circumstances, in this way augmenting yield potential. The consolidation of reenacted strengthening adds a modern layer to the dynamic cycle in horticulture, adjusting planting times with ideal climatic and occasional circumstances.

2. LITERATURE SURVEY

Pudumalar, S., Ramanujam, E., Rajashree, R. H., Kavya, C., Kiruthika, T., & Nisha, J. collaborated to create a web-based farming assistance application [4]. This application incorporates a secure login and registration system utilizing OTP for enhanced authentication, mitigating potential password-based vulnerabilities. The primary goal of the application is to facilitate direct farmer-to-dealer interactions, promoting greater profitability for farmers. Notifications about new advertisements, predictions for profitable product selling in different states or locations, and real-time payment options through UPI and credit cards are key features. The developers employed a system based on the XGBoost algorithm, utilizing a Decision Tree Classifier to forecast soil fertility and identify suitable crops for specific conditions. Additionally, their model suggests appropriate fertilizers to replenish lost nutrients. The application takes input parameters such as nitrogen (N), phosphorus (P), potassium (K), temperature, humidity, pH, and rainfall to provide crop recommendations based on the XGBoost model.

Sheng-Feng Kuo, Chen-Wuing Liu, Gary P.Merkley developed a research on agricultural water resource water management[5]. The primary focus of their research is the development of a model for on-farm irrigation systems, aiming to optimize the allocation of irrigated areas for different crops to maximize net benefits. They employ a customized simulated annealing method to achieve this goal. The model is designed based on on-farm irrigation scheduling and simulated annealing to enhance planning and management of agricultural water resources. Simulated annealing, a method recently applied to functional optimization problems, addresses 'real-world' issues where the objective is to attain minimum or maximum global values within specific constraints. In the context of irrigation project planning, the 'real-world' problem involves determining optimal crop area allocations to maximize the benefits of the project while considering various constraints. The researchers apply their model to Delta, Utah, using simulated annealing to optimize crop production benefits and determine crop area allocations. Unlike traditional optimization methods, simulated annealing's ability to update configurations, even without energy improvement from the previous simulation 'temperature,' helps overcome the problem of getting stuck in local optima. Traditional methods only update configurations when there is an improvement in energy from the previous iteration, limiting their effectiveness.

Andre Gloria, Joao Cardoso, and Pedro Sebastiao devised a system employing a diverse array of strategically positioned sensors across agricultural fields to gather essential data for accurate monitoring[6]. The data is transmitted via a Wireless Sensor Network (WSN) using NB-IoT to a cloud server, where it is stored and subjected to analysis through machine learning techniques. The objective is to determine optimal actions for field management. Their research also delves into identifying the most effective algorithm for irrigation scheduling based on sensor data, comparing the performance of machine learning approaches against traditional or smart irrigation solutions. The supervised learning aspect is categorized into classification and regression. Classification methods aim to approximate a mapping function from input data to identify output values. The study concluded that Random Forest emerged as the most effective solution, achieving an impressive accuracy of 84.6%. Notably, Random Forest outperformed SVM and Decision Trees by nearly 7%, with Neural Network being the only technique exhibiting results closely resembling Random Forest, albeit with only a 4% difference.

Emmanuel Abiodun Abioye, Oliver Hensel, Travis J. Esau, Olakunle Elijah, and Mohamad Shukri Zainal Abidin have proposed the development of wireless sensor network (WSN) technologies tailored for intelligent agricultural applications through remote sensing[7]. The controlled monitoring of agricultural processes has facilitated a deeper understanding of the dynamic changes in weather, soil, and crop conditions during the growing season. The study implemented an intelligent irrigation management approach with remote monitoring, utilizing the KNN model to classify crops based on water requirements and drought sensitivity in various regions. The on-and-off pump motor was activated using a float sensor. The study's findings suggest that sustainable precision irrigation management plays a crucial role in achieving food security and preventing water scarcity. Additionally, the paper extends the exploration of machine learning techniques applied to irrigation management, including supervised, unsupervised, and reinforcement learning. The research indicates that the selection of a machine learning model for irrigation management depends on factors such as the availability of experimental data sets, computational complexity, the nature of implementation, and deployment type. The paper discusses challenges and opportunities in the application of machine learning techniques and digital solutions. Notably, supervised and unsupervised learning have shown positive outcomes in precision irrigation based on the reviewed findings[8].

3. PROPOSED METHOD & RESULTS

To enhance farmers' income by determining the most suitable crop to grow during specific periods, a research study suggests integrating two highly effective computational techniques, namely xgboost and simulated annealing. By combining these methods, farmers can anticipate and select the crops that would maximize their profits.

3.1 Library collection and Data Preprocessing:

For this study aimed toward improving farmers` earnings through specific crop selection, a complete library series and meticulous records preprocessing are essential components. The datasets applied had been sourced from Kaggle, encompassing precious records on climate patterns, numerous soil kinds, and crop records throughout numerous areas in India.

3.2 Library Collection:

To cope with the intricacies of records manipulation, predictive modelling, and optimization, a set of effective libraries became amassed. The number one libraries enlisted consist of Pandas for green records dealing with and manipulation, Scikit-Learn for device gaining knowledge of functionalities, Matplotlib and Seaborn for records visualization, XGBoost for predictive modelling, and SciPy for enforcing the Simulated Annealing optimization set of rules. These libraries together empower the observer with a sturdy set of gear to extract insights, educate models, and optimize crop schedules effectively.

3.3 Data Preprocessing:

The gathered datasets underwent meticulous preprocessing to ensure suitability for analysis, concerning dealing with lacking values, encoding specific variables, and scaling numerical functions with the use of Scikit-Learn's functionalities. This aimed to create a clean, standardized dataset conducive to correct predictive modelling and optimization. The Kaggle datasets supplied a wealth of records critical for the observation, encompassing beyond climate records for weather insights, soil kinds for agricultural panorama details, and crop records with historic overall performance metrics. This numerous datasets served as the inspiration for the education of the XGBoost model, which, in turn, was knowledgeable about the Simulated Annealing set of rules for optimizing crop schedules. Place table titles above the tables.

3.4 Exploratory Data Analysis:

EDA is carried out as preparatory step before undergoing the analytical analysis or modelling.This research aim to represent the prominent characteristics, find trends, track outliers, and determine how variables relate to one another.

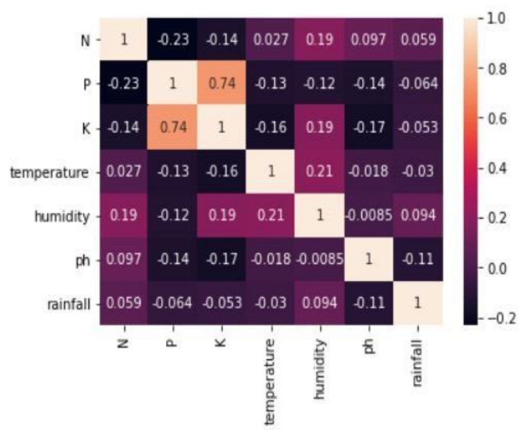


Figure 1: Correlation between variables.

The above heat map depicts how the variables correlated to each other and how one factor affects the other. If the result is less than 1 and greater than zero, then it is a positive correlation. And also there are negative values the existence of those values indicates inversely proportional to each other.

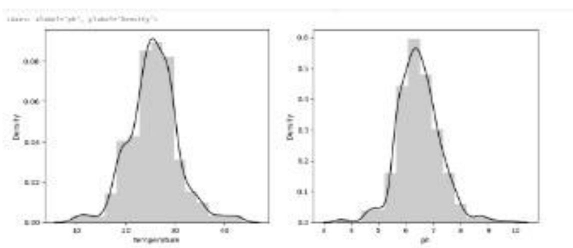


Figure 2: Distribution of temperature and ph.

It Depicts that The bell-shaped and symmetrical data indicates that while most trials get results close to the mean, there are sometimes significant deviations. It's also intriguing to see how closely these two resemble one other.

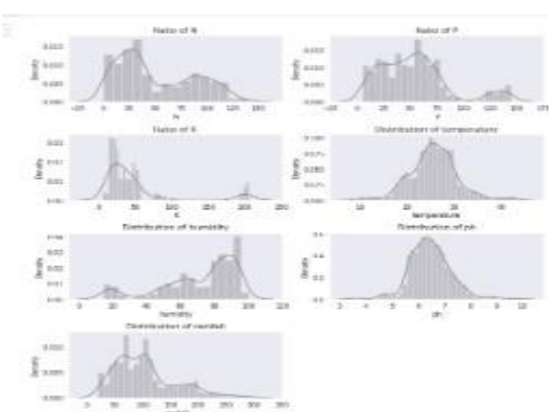


Figure 3: Distribution plot for distinct characteristics.

In the above image illustrates the Overall distribution of various characteristics like n, k, temperature, humidity and rainfall. Distribution. plots visually analyze the distribution of sample data by comparing its actual distribution to the theoretical values anticipated from a given distribution.

3.5 Implementation

3.5.1 XGBoost

This targets to are expecting the yield of plants in precise weather situations the use of facts gathered from the authority's internet site www.facts.gov.in for the years 20002014. The gathered facts consist of most temperature, minimal temperature, season-smart rainfall, region of cultivated land, and manufacturing of rice. When carried out to 4 algorithms—Linear Regression, Decision Tree Regression, Random Forest Algorithm, and XG Boost Algorithm—to the gathered datasets to assess their accuracy in predicting crop yield. The assessment metrics used are R^2 (coefficient of determination) and Mean Square Error (MSE).that XG Boost Algorithm executed a great many of the 4 algorithms in predicting crop yield.

Additionally, the contrast of Mean Square Error (MSE) values additionally helps the realization that XG Boost Algorithm is the great-appearing set of rules. The MSE values for the algorithms are as follows:

- Linear Regression: 2972466269.2235
- Decision Tree Algorithm: 2847222194.9242
- XG Boost Algorithm: 1999378847.4864

The decreased MSE cost for XG Boost Algorithm shows better accuracy and perfection in prediction.primarily based totally on each R^2 and MSE comparison, indicates that the XG Boost Algorithm is the only set of rules for predicting crop yield inside the given weather situations and dataset.

Step 1: Compute the Residuals and Make an Initial

Prediction

Step 2 is to construct an XGBoost tree.

Step 3 is to prune the tree.

Step 4 is to determine the output values of the leaves.

Step 5: Form New Forecasts

Step 6: Utilizing the New Predictions, Calculate Residuals

Step 7: Repetition of Steps 2–6

In this method, the algorithm determines which crop to grow in the given location.

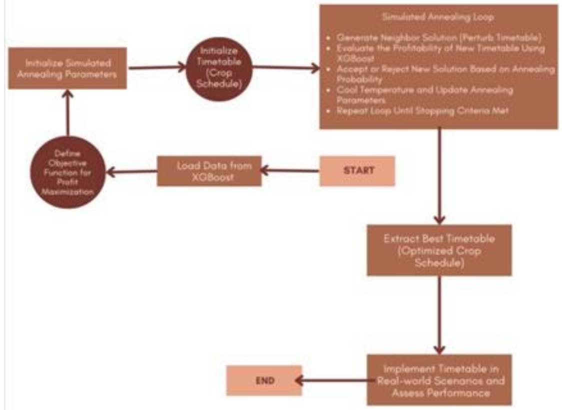
3.5.2 Stimulated Annealing:

Simulated annealing is a probabilistic method for approximating the worldwide most beneficial of a given function. Specifically, it's miles a metaheuristic to approximate worldwide optimization in a huge seek area for an optimization problem. For huge numbers of nearby optima, SA can discover the worldwide optima.

Simulated Annealing (SA) is a computational approach used for purposeful optimization troubles in irrigation assignment planning. Its objective is to discover minimal or most worldwide values inside distinct constraints, maximizing crop place allocation. The SA module starts evolving via way of means of receiving output from the irrigation scheduling module, then inputs simulated annealing parameters through a personal interface. A design "chromosome" is defined, generated, and decoded right into an actual number. Constraints are implemented to make certain the answer adheres to limitations. A goal feature and health price are implemented to assess the answer`s quality. The annealing agenda is carried out with the use of Boltzmann probability, and parameters like cooling price and termination criterion are set.

The flowchart is used to integrate Simulated Annealing with XGBoost to optimize a yearly timetable for crop cultivation.

The simulated annealing loop's iterative structure makes it possible to experiment with various crop plans while taking XGBoost recommendations and profitability goals into account. Real-world agricultural scenarios can be used to apply and evaluate the final optimized schedule.



4. RESULT

The studies has the evaluation of 4 algorithms—Linear Regression, Decision Tree Regression, Random Forest Algorithm, and XGBoost Algorithm—at the dataset from www.facts.gov.in (years 2000-2014) and kaggel for predicting crop yield discovered specific overall performance variations. Evaluation metrics, in particular Mean Square Error (MSE), have been hired to gauge accuracy. Linear Regression and Decision Tree fashions displayed notably excessive MSE values, suggesting suboptimal overall performance. However, the XGBoost Algorithm exhibited a extensively decrease MSE (1999378847.4864), indicating advanced accuracy in predicting crop yield. This end turned into in addition supported through comparisons of each R^2 and MSE metrics. The stepwise XGBoost methodology, related to residual computation, tree construction, pruning, and iterative refinement, showcased its effectiveness in figuring out the maximum appropriate vegetation for cultivation in particular climate conditions, improving its desire for this predictive task.

5. CONCLUSION

To summarize, the suggested model has followed a methodical process that includes exploratory data analysis, calculated data partitioning, Simulated annealing, XGBoost algorithm application, and careful hyperparameter tuning. During these phases, the study methodically tackled the difficulties by figuring out which crops, based on several factors, are most suited for a certain place and when to cultivate particular crops. The trip produced an improved XGBoost model with notable improvements in recall and precision. The model's enhanced performance in identifying the types of crops to be grown was demonstrated by this all encompassing approach, which combined machine learning techniques with strategic data augmentation. It underscored the efficacy of these methods in optimizing crop choices and

furnished farmers with a timetable to ascertain viable crops for specific times in the agricultural calendar. The research provides a unified framework, seamlessly combining various methodologies, offering a robust solution to enhance the accuracy and efficiency of crop production systems.

VI. FUTURE SCOPE

Considering prospects for the future, the suggested study establishes the groundwork for improvements in crop production methods. Extending the system's reach to remote villages is crucial, promoting inclusive agricultural practices and aiding farmers in underserved areas. Furthermore, the system could evolve to predict not only the types of crops suitable for specific regions but also anticipate potential crop diseases. Integrating information on fertilizers and providing comprehensive agricultural insights could contribute significantly to sustainable farming practices and empower farmers with holistic guidance for optimal crop cultivation. Using some data visualization tools such as tableau and power bi an interactive dashboard can be developed to provide users with an intuitive and user friendly interface, facilitating seamless interaction and data interpretation. These developments, which combine farming productivity gains with technological improvements, have the potential to completely transform the agricultural landscape in the future.

The contents of the journal are peer-reviewed and archival.

REFERENCES

[1] Odegard, I. Y. R., & Van der Voet, E. (2014). The future of food—Scenarios and the effect on natural resource use in agriculture in 2050. Ecological Economics, 97, 51-59.

[2] Ederer, P., Baltenweck, I., Blignaut, J. N., Moretti, C., & Tarawali, S. (2023). Affordability of meat for global consumers and the need to sustain investment capacity for livestock farmers. Animal Frontiers, 13(2), 45-60.

[3] Zeferino, J. A., Antunes, A. P., & Cunha, M. C. (2009). An efficient simulated annealing algorithm for regional wastewater system planning. Computer-Aided Civil and Infrastructure Engineering, 24(5), 359-370.

[4] Pudumalar, S., Ramanujam, E., Rajashree, R. H., Kavya, C., Kiruthika, T., & Nisha, J. (2017, January). Crop recommendation system for precision agriculture. In 2016 Eighth International Conference on Advanced Computing (ICoAC) (pp. 32-36). IEEE.

[5] Kuo, S. F., Liu, C. W., & Merkley, G. P. (2001). SW— Soil and water: application of the simulated annealing method to agricultural water resource management. Journal of Agricultural Engineering Research, 80(1), 109-124.

[6] Glória, A., Cardoso, J., & Sebastião, P. (2021). Sustainable irrigation system for farming supported by machine learning and real-time sensor data. Sensors, 21(9), 3079.

[7] Bassine, F. Z., Epule, T. E., Kechchour, A., & Chehbouni, A. (2023). Recent applications of machine learning, remote sensing, and iot approaches in yield prediction: a critical review. arXiv preprint arXiv:2306.04566.

[8] Madhukar Rao, G., & Ramesh, D. (2016). Supervised learning techniques for big data: a survey. IJCTA. Int Sci Press, 9, 3811-3891.

[9] Shashikant, V., Sahu, R. T., Prasad, A. D., & Verma, M. K. (2022, May). Development of an optimal operating policy of multi-reservoir systems in Mahanadi Reservoir Project Complex, Chhattisgarh. In Journal of Physics: Conference Series (Vol. 2273, No. 1, p. 012020). IOP Publishing.