Optimization of crop production: A Hybrid Technique Using Simulated Annealing and XGBoost for Crop Profit Maximization

Mallyala Nissie[1], Bellam Keerthana[2], Perumalla Mounika Reddy[3], Mekala Sravani[4], Amrutha Varshini[5], G Madhukar Rao[6]

[1]Department of Computer Science and Engineering Koneru Lakshmaiah Education Foundation
Hyderabad India-500075
nissiehft@gmail.com

[2]Department of Computer Science and Engineering Koneru Lakshmaiah Education Foundation
Hyderabad India-500075
keerthanabellam2310@gmail.com

[3]Department of Computer Science and Engineering Koneru Lakshmaiah Education Foundation
Hyderabad India-500075
Mounikareddy1459@gmail.com

[4]Department of Computer Science and Engineering Koneru Lakshmaiah Education Foundation
Hyderabad India-500075
Sravanichowdary1459@gmail.com

[5]Department of Computer Science and Engineering Koneru Lakshmaiah Education Foundation
Hyderabad India-500075
amruthavarshini9966@gmail.com

[6]Department of Computer Science and Engineering Koneru Lakshmaiah Education Foundation
Hyderabad India-500075
madhusw511@klh.edu.in

ABSTRACT:

This study investigates the use of a novel hybrid technique to optimize crop production in various regions of India. By effectively managing different agricultural parameters, the proposed approach integrates XGBoost and Simulated Annealing with the goal of maximizing crop profit. In order to optimize resource allocation and crop planning, Simulated Annealing is used to fine-tune the decision variables. Meanwhile, XGBoost, a potent machine learning algorithm, helps with predictive modeling for yield estimation. Combining these methods provides a thorough approach to improving agricultural output while accounting for the intricate and ever-changing cropgrowing landscape of India. The outcomes show how well the hybrid technique works to maximize crop profits, offering important new information for sustainable and profitable agricultural methods in Indian context

Keywords : Precision farming; XGBoost,; Sustainable agriculture; Simulated annealing; Climate adaptation.

1. INTRODUCTION

The foundation of our society, farming nourishes billions of people globally and is their primary source of income. Ancient people were able to adjust to their needs since they harvested food on their land. Therefore, a variety of species, including humans, livestock, and birds, grow and use natural crops. There is a growing need for creative solutions in the agriculture sector due to factors involving a population boom, an erratic climate, and the requirement for sustainable methods[1].There are serious issues facing the world food system now, and these issues should only get worse over the next forty years. If there is enough will and funding, advancement may be made quickly with current technologies and knowledge. To find creative answers, tackling upcoming problems will necessitate more significant research funding and drastic adjustments. Deterioration of ecosystems, exhaustion of natural resources, stagnating farm earnings, fragmented land holdings, and a lack of jobs in non-farm industries are some of the problems that exist today. Using new technologies is thought to be a crucial approach to raising agricultural productivity in the future. Rather than managing entire fields based on hypothetical averages, precision farming recognizes the distinctions between each site and modifies management strategies accordingly.The research being investigated explores an innovative approach to provide farmers with practical knowledge about which crops are best to grow given particular soil types and other critical characteristics. Its goal is to create an intelligent system that looks at important factors and creates a dynamic schedule that gives farmers specific advice on what crops are best for different regions and times of year. This approach utilizes modern analytics to deliver customized and effective farming schedules to improve decision-making for sustainable and fruitful agricultural operations. The approach begins with a careful focus on data quality since reliable and precise information serves as the foundation for all analytical processes. A thorough phase of data preparation and cleaning ensures the integrity of the insights obtained by laying the groundwork for further studies. An exploratory data analysis (EDA) is conducted based on the crop suggestion dataset to suggest crops that may be produced in suitable circumstances while taking into account certain soil parameters.

The strategic integration lies at the heart of our research. With the use of XGBoost, we can determine which crops are most suited for a certain area depending on several variables. This improves crop selection precision, maximizing agricultural techniques for the specific region and fostering effective and sustainable farming. A robust machine learning system called XGBoost examines a wide range of variables, including soil type, climate, and past data, to deliver precise recommendations for the best crops to grow in a particular region. Through the use of this cutting-edge predictive modeling, farmers can make well-informed decisions on sustainable and productive agricultural methods by gaining a thorough grasp of crop compatibility. This creative strategy promotes greater production and endurance in the face of changing difficulties in the agricultural industry, representing a major step towards precision farming[2].

To determine the optimal time for cultivating specific crops, the application of simulated annealing emerges as a valuable solution. Simulated annealing, a metaheuristic algorithm inspired by annealing processes in metallurgy, contributes to the optimization of planting schedules. By systematically exploring various temporal possibilities and adapting to changing conditions, simulated annealing aids in identifying the most favorable periods for crop cultivation.This approach enhances precision in agricultural planning, allowing farmers to synchronize planting with environmental conditions, thereby maximizing yield potential. The incorporation of simulated annealing adds a sophisticated layer to the decision-making process in agriculture, aligning planting times with optimal climatic and seasonal conditions.

2. LITERATURE SURVEY

Pudumalar, S., Ramanujam, E., Rajashree, R. H., Kavya, C., Kiruthika, T., & Nisha, J. collaborated to create a webbased 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**.**

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.

3.1 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.

3.2 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.

3.3 Implementation

3.3.1 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 2000-2014. 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:

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.

3.3.2 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.

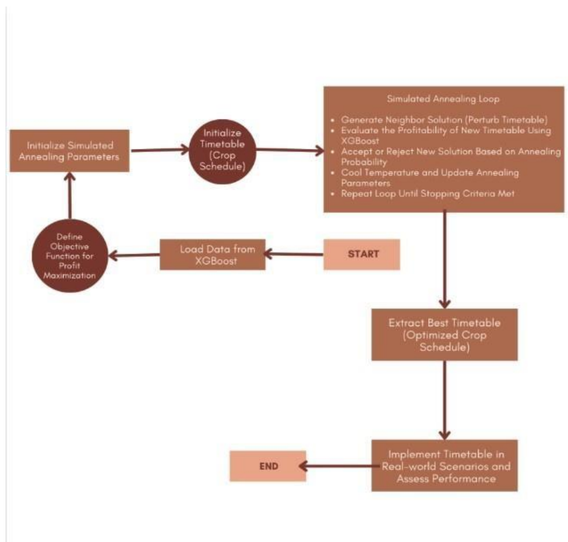

Figure 3: Flowchart of Algorithm implementation.

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.

Table 1 Comparison of algorithms

Algorithm	**Accuracy**
Linear Regression	2972466269.2235
Decision Tree Algorithm	2847222194.9242
XG Boost Algorithm	1999378847.4864

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.

6. 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.

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] 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.

[5] 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.
[6] 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.
[7] Madhukar Rao, G., & Ramesh, D. (2016). Supervised learning techniques for big data: a survey. IJCTA. Int Sci Press, 9, 3811-3891.