AGRICULTURAL CROP YIELD PREDICTION

USING AI

ARUNKUMAR A

Department of Information Technology

Sri Sivasubramaniya Nadar College of Engineering Tamil Nadu [arun2110763@ssn.edu.in](mailto:arun2110763@ssn.edu.in)

ASHWIN M

Department of Information Technology

Sri Sivasubramaniya Nadar College of Engineering Tamil Nadu [ashwin2110586@ssn.edu.in](mailto:ashwin2110586@ssn.edu.in)

BADRI NARAYANAN V T

Department of Information Technology

Sri Sivasubramaniya Nadar College of Engineering Tamil Nadu [badri2110867@ssn.edu.in](mailto:badri2110867@ssn.edu.in)

Dr. N. Radha Dr. R. Swathika

Department of Information Department of Information

Technology Technology

Sri Sivasubramaniya Nadar Sri Sivasubramaniya Nadar

College of Engineering College of Engineering

Tamil Nadu. Tamil Nadu.

*Abstract* - Agriculture plays a pivotal role in India’s economic growth, influenced by diverse factors such as market dynamics, production rates, soil conditions, and climatic variations. This research employs Constraint Satisfaction Problem (CSP) and Naive Bayes algorithms to predict crops and suggest fertilizers, with the primary aim of optimizing production and increasing profits for farmers. Utilizing a dataset sourced from an Indian government website, comprising over a thousand records filtered using Python Pandas, the study investigates the potential benefits of integrating AI algorithms into modern agriculture. The focus is on providing farmers with informed decisions to enhance economic returns and ensure a steady supply to meet consumer needs. Crop ranking, determined by the degree of constraint fulfilment, is central to the methodology, with the top- ranked crop satisfying most of the specified conditions. The envisioned outcome is a groundbreaking software solution poised to elevate farmers' profits and efficiently address current market demands in the agricultural industry. Moreover, the incorporation of additional features aims to assist farmers in disease detection, adding an extra layer of practicality to the proposed solution

***Keywords: Crop Prediction, Artificial Intelligence, CSP, Naive Bayes, Fertilizer Suggestion, Indian agricultural dataset***

# INTRODUCTION

Agriculture is a vital element that has a significant role in nourishing the world’s growing population. To keep pace with the increasing demand for foodstuffs, farmers need to make the best use of them to reap output while minimizing losses. Forecasting and examining reap growth is a serious part of modern agriculture, and Artificial Intelligence and Machine Learning has become a powerful tool to achieve this goal line

In recent years, the global agricultural landscape has witnessed unprecedented challenges arising from climate change, population growth, and resource constraints. As a result, the need for precision agriculture and data-driven decision-making has become increasingly imperative to ensure food security and sustainable agricultural practices. One key aspect of this paradigm shift is the development and implementation of robust crop yield prediction systems, which serve as invaluable tools for farmers, policymakers, and researchers alike.

Accurate and timely prediction of crop yields plays a pivotal role in optimizing resource allocation, mitigating production risks, and fostering sustainable agricultural practices. Traditional methods of yield estimation often rely on historical data and manual observations, falling short in providing real-time insights and adaptability to environmental conditions. The integration of cutting-edge technologies, such as machine learning, AI Algorithms and data analytics, open new avenues for revolutionizing crop yield prediction and addressing the intricacies of modern agricultural challenges.

This research paper seeks to explore and contribute to the evolving field of agricultural crop yield prediction by presenting a comprehensive review of existing methodologies, challenges, and opportunities. The significance of this research lies in its potential to empower farmers with timely and accurate information, enabling them to make informed decisions about crop management practices, resource allocation, and risk mitigation. Furthermore, the proposed framework contributes to the broader discourse on sustainable agriculture, aligning with global initiatives to achieve food security while minimizing environmental impact.

The environmental data, soil data, and other data collected from trusted websites are gathered and processed to ensure that they don't have any inappropriate values. After knowing certain details about the land and its characteristics, users can make use of this system, which utilizes CSP along with Naive Bayes, to rank the top three crops that are suitable for their land. The system considers factors such as market demand and profitability for both the farmer and the customer. The aim is for farmers to use these technologies to achieve their goal of improved harvests by making better selections for crops to be cultivated in their field.

# LITERATURE SURVEY

In [1], the authors have implemented a system that uses machine learning techniques to predict the most suitable crop type from the surrounding environmental factors for plant growth and processing them with the trained sub-models of the main system. They used machine learning techniques such as Naive Bayes, Support Vector Machine, K- Means Clustering and also Natural Language Processing (Sentiment Analysis) concerned with Artificial Intelligence to recommend a crop for the selected land with the site-specific parameters with high accuracy and efficiency.

In [2], the potential benefits of integrating machine learning algorithms in modern agriculture are investigated. The main focus of these algorithms is to help optimize crop production and reduce waste through informed decisions regarding planting, watering, and harvesting crops. In this research paper, fifteen different algorithms have been considered to evaluate the most appropriate algorithms to use in agriculture, and a new feature combination scheme-enhanced algorithm is presented. The results show that it is possible to achieve a classification accuracy of 99.46% using Naïve Bayes Classifier and Hoeffding Tree algorithms. These results will indicate an increase in production rates and reduce the effective cost for the farms, leading to more resilient infrastructure and sustainable environments.

In [3], the planned work introduces an efficient degree economical crop recommendation system. Use of naïve mathematicians makes the model terribly economical in terms of computation. The system is scalable because it may be worth taking a look at on totally different crops. From the yield graphs the simplest time of sowing, plant growth and gathering of plants may be known. Conjointly the best and worst condition may also be incurred. The model focuses on all styles of farms, and smaller farmers may also be benefitted. This model may be increased to seek out the yield of each crop, and for chemical recommendation. Conjointly it may be changed to recommend concerning the fertilizers and irrigation want of crops.

In [4], the authors evaluated the performance of the GNB algorithm when combined with different feature scaling and feature extraction techniques in stock price movement prediction. GNB produced better performance with Min-Max scaling technique than it does with standardization scaling techniques. With this research paper we can collect some ideas on how to apply Gaussian Naive Bayes Algorithm along with Minmax Scalar to predict crops.

In [5], many research papers (about 500 papers) on crop prediction were evaluated out of which temperature, rainfall and soil type were found to be the best parameters which is used to know the parameters that are required to predict the crop. In [6] the authors have used a color-based image classification, leveraging deep convolutional neural networks, to predict plant’s disease. The study, using ResNet-50, achieved an impressive 96.35% accuracy in classifying tomato plant diseases, offering a promising solution. Conventional rice plant disease diagnosis is subjective and time- consuming, impacting productivity. This research paper is useful in developing a system to predict diseases in plants.

In [7], the basics of the Naive Bayes algorithm are explained with the help of 6 steps, using which we were able to correlate it with the data we have. In [8], the author used big dataset which consist of more than two lakhs of records to predict the crop yield based on big data analysis technology, which differs in traditional method in the structure of handling and in the means of modelling. With help of this we can to know that large number of records will help to predict more accurate and precise crop.

With the help of [9] we came to know that we can use multiple machine learning models such as KNN, Decision tree, Regression, SVM, Random Forest, Naive Bayes. ANN and Deep Neural Networks can also use to predict the crop in which except SVM all other models are able to predict the crop with more 93% accuracy. In [10], the authors have concluded that using machine learning models like K- NN and Naive Bayes the crop can be predicted more accurately. Naive Bayes has accuracy of 91% and KNN with 75%. With help of this, conclusion is made that NB can also be part of our system.

In [11], the authors used BPN (back propagation neural networks) and API (application user interface) which is used to get input such as weather based on which temperature is calculated. In [12], the author has concluded that using Naive Bayes along with supervised learning will improve the success rate of crop prediction and also increases the effective farming.

In [13], the authors describes the development of a different crop yield prediction model with ANN, with 3 Layer Neural Network. Initially the result obtained considering optimizer RMS prop with accuracy 45 %, later it will be enhanced to 90% by increasing layers, adjusting weight, bias and changing optimizer to Adam.

In [14], the authors have implemented various machine learning techniques to estimate the crop yield in Rajasthan state of India on five identified crops. The results indicate that among all the applied algorithms; Random Forest, SVM, Gradient Descent, long short-term memory, and Lasso regression techniques; the random forest performed better than others with 0.963 R2, 0.035 RMSE, and 0.0251 MAE. The results were validated using R2, root mean squared error, and the mean absolute error to cross-validation techniques. This paper intends to put the crop selection method into practice to help farmers solve crop yield problems.

# METHODOLOGY

1. *Dataset*

It is required to have a specific range of temperature, humidity, soil pH, sunlight, and soil moisture for a plant to be grown healthy. To receive a good harvest these conditions need to be satisfied but it may vary according to the plant varieties. The initial set of data is collected from various trusted agricultural websites and in the absence of many records, synthetic data is used. Initial set of datasets is used to check whether the prediction based on the CSP algorithm is accurate.

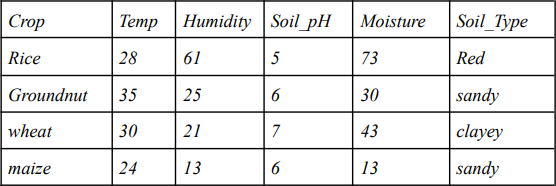


Table 1: Sample Data

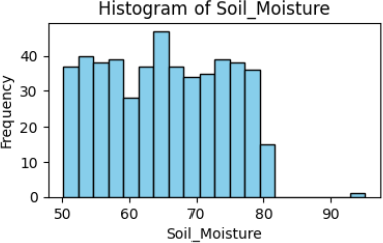


Fig 1: Distribution of Soil moisture

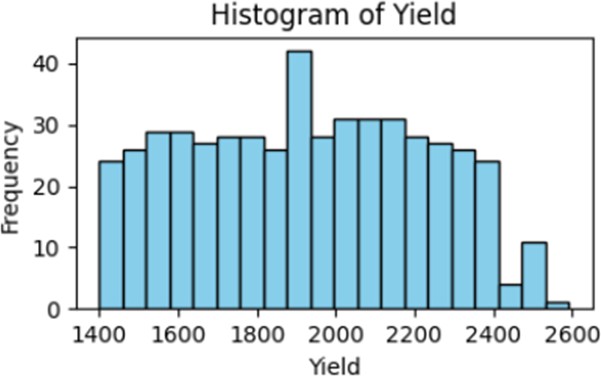


Fig 2: Frequency of Yield

1. *Constraint Satisfaction Problem*

The Constraint Satisfaction Problem (CSP) is a formalism used in Artificial Intelligence (AI) to model and solve problems that involve a set of objects whose state must satisfy several constraints. These problems are pervasive across various domains, including scheduling, resource allocation, configuration, and

planning. The primary goal of a CSP is to find a consistent assignment of values to a set of variables that satisfies a given set of constraints.

Key components of a CSP:

Variables (X): These represent the objects or entities in the problem that need to be assigned values. For

example, in a scheduling problem, variables could represent different tasks or events.

Domains (D): Each variable has a domain, which is the set of possible values it can take. The domains represent the feasible options for each variable. In a scheduling problem, the domain for a variable could be the set of possible time slots for a task.

Constraints (C): Constraints define relationships or restrictions on the combinations of values that can be assigned to variables. Constraints specify which combinations of values are allowed and which are not. They capture the rules and requirements of the problem. For instance, a constraint in a scheduling problem might state that two tasks cannot occur simultaneously.

The primary task in solving a CSP is to find a consistent assignment of values to variables that satisfies all the constraints. This solution is often referred to as a "solution" or "assignment." The process of solving a CSP involves searching through the space of possible assignments systematically, taking into account the constraints to eliminate inconsistent options.

In our project, we have chosen the domain as the different types of crops and the constraints as the inputs which are rendered from the user/farmer. Finally, the variables hold the solution which satisfies the constraints and provides a solution within the domain using backtracking. The algorithm ranks the top three most suitable crops which can be cultivated in their land along with the necessary details about that crops.

1. *Gaussian Naive Bayes Algorithm*

The Gaussian Naive Bayes algorithm is a classification algorithm based on Bayes' theorem, which is a probabilistic approach for making predictions. It is particularly well-suited for classification tasks where the features are continuous and assumed to follow a Gaussian (normal) distribution. The "naive" in its name comes from the assumption that features are conditionally independent given the class label, which simplifies the modelling and computation but might not always hold true in real- world scenarios.

We have used the Gaussian Naive Bayes algorithm in our project to calculate the percentage probability of growth of the chosen top three crops which are most suitable for cultivation based on their inputs. It calculates the success probability of each crop so that more is gained by the farmer on what crop should be cultivated in their land.

Formula:

Bayes theorem allows to calculate posterior probability P(C|X) from the given P(X|C), P(X) and P(C).

P(C|X) = conditional probability of X when given C that is the posterior probability.

P(X|C) = conditional probability of C when given X that is likelihood. P(C) = prior probability of C.

P(X) = probability of X Naive Bayes Classifier: Formula -

* n is number of training examples for which =
* is number of examples for which = and .
* is prior estimation for .
* is equivalent sample size.

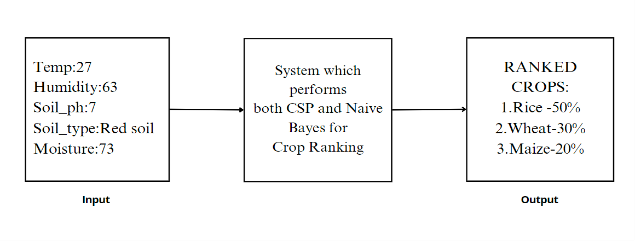


Fig 3: Sample output given to Farmers

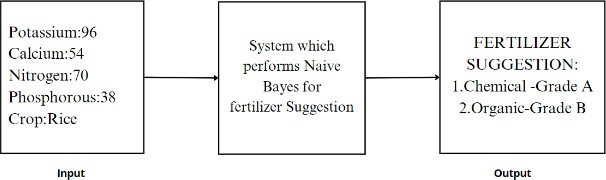
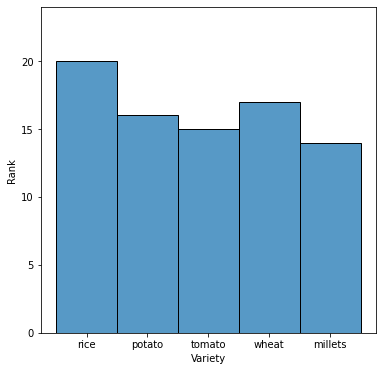


Fig 4 : Sample output for suggesting Fertilizer

# .

# EXPERIMENT

Using CSP and Naive Bayes algorithms the crops are predicted that satisfies most of the constraints and also Naive Bayes predict the crop with accuracy of 95%.The Accuracy of predicting the crop increases when more data is provided. The Agent takes various inputs as mentioned above. There are some limitations with those inputs. For example: the Soil PH value must be in the range of 4 to 13. As the data source contains outliers it is handled in by filling it with mean value and if it has null value then it is removed.



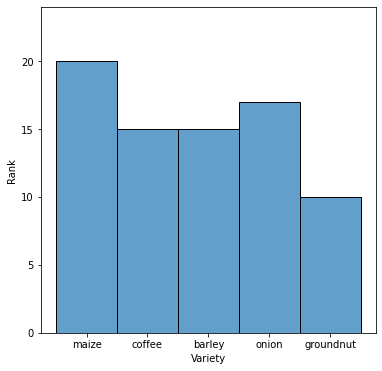


Fig 5: Ranking of Crops

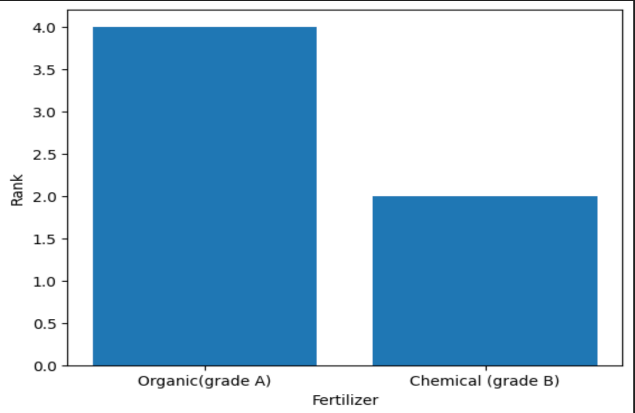


Fig 6: Fertilizer Suggestion

Based on the input given by the farmers it ranks the crops. In which the top 3 ranked crops will be displayed as output. In Fig 5 rice has the highest rank as its conditions needed for growing are mostly satisfied and followed by millet and coffee. In Fig 6 it shows what type of fertilizer can be used for the specified crop. For example if the user input for fertilizer suggestion is for the crop rice then it shows that you can use organic grade A fertilizer if it not possible then try to use chemical grade B fertilizer.

Based on the input provided and data collected, the system can suggest the most suitable crop and probability percentage and to the farmers which increases their profit and also increases their knowledge of cultivation of crops in the future. The fig.3 shows the sample result where the farmers get ,once the input for all the factors are entered. By providing large amounts of data the accuracy of the prediction of the crop is also increased. The more data the system feeds, the more accurate it will become. On integrating it with any machine learning concepts we can also check the accuracy of the system. The fig.4 shows the sample result the farmers get once the input about their soil nutrients are given. It also gives the description for choosing the particular fertilizer. For example if it recommends Chemical fertilizer with grade A then the description will be as the nutrient content like potassium is below the needed so it can be increased by adding the chemical fertilizer of grade A if not you can also use organic fertilizer of grade B.

# CONCLUSION:

In a modern environment with less space and less knowledge of agriculture, all the factors are considered from the perspective of farmer and plant, and the farmer is properly guided until the harvesting. Before selecting any plant to grow it is important to have the knowledge and an understanding of the factors that affect the cultivation and how to maintain or control them. From this system, these above-mentioned factors are automatically processed and provide the crop type most suitable for cultivation.

This system also calculates the market demand, yield and quantity required for the crops recommended. So the farmers will gain more and more knowledge about cultivating practices, which crop is most suitable under which climatic conditions which in turn increases their profitability. This system will definitely be a torch bearer for future agricultural systems and for the farmers.

# FUTURE WORKS

It includes more features to gather feedback from farmers and make recommendations for crops. With the help of this feedback, we can make changes to enhance the performance of the system. Additionally, we are developing an app that will work on any type of smartphone. Currently, we collect detailed information about the land from the user. As an additional feature, we plan to implement an algorithm to predict certain factors using other relevant factors. For example, we could predict soil nutrition by analysing soil type and pH level. This enhancement will make the system more user-friendly for farmers.

# REFERENCES

1. Bandara, Weerasooriya, Ruchirawya, Nanayakkara, Dimantha, Pabasara (2020). Crop Recommendation System.International Journal of Computer Applications (0975 – 8887). Volume 175– No. 22.
2. Elbasi, Zaki,Topcu, Abdelbaki, Zreikat, Cina, Shdefat, Saker(2023).

Crop Prediction Model Using Machine Learning Algorithms. Applied sciences.

1. Priya, Ramesh, Khosla(2018). Crop Prediction on the Region Belts of India: A Naïve Bayes MapReduce Precision Agricultural Model.

2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)

1. Ampomah, Nyame, Qin, Addo, Gyamfi, Micheal. Stock Market Prediction with Gaussian Naïve Bayes Machine Learning Algorithm. Informatica 45 (2021)

243–256 243

1. Klompenburg, Kasahun,Catal. Crop yield prediction using machine learning: A systematic literature review. Computers and Electronics in Agriculture 177 (2020)

105. Computers and Electronics in Agriculture 177 (2020) 105.

[6]U.Archana, Amanulla Khan, Appani Sudarshanam, C.Sathya, Ashok Kumar Koshariya,R.Krishnamoorthy(2023).Plant Disease Detection using ResNet. Proceedings of the International Conference on Inventive Computation Technologies

# (ICICT 2023).

1. Sunil Ray (2015), ‘6 Easy Steps to Learn Naive Bayes Algorithm with codes in Python and R’, Analytics Vidhya.
2. Anita Gehlot, Dr. Neeru Sidana, Deepali Jawale,Dr.Neetu Jain,Bhaskar Pratap Singh and Dr. Barinderjit Singh(2022).Technical analysis of crop production prediction using Machine Learning and Deep Learning Algorithms. International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES) | 978-1-6654- 7413-9/22/$31.00 ©2022 IEEE | DOI:

# 10.1109/ICSES55317.2022.9914206

1. Ramesh Medar&Anand M. Ambekar, “Sugarcane Crop prediction Using Supervised Machine Learning", International Journal of Intelligent Systems and Applications Volume: 3 | August 2019.
2. R. Medar, V. S. Rajpurohit and S. Shweta, "Crop Yield Prediction using Machine Learning Techniques," 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), 2019
3. Sachee Nene &Priya, R “Prediction of Crop yield using Machine Learning”, International Research Journal of Engineering and Technology (IRJET)

Volume: 05 Issue: 02 | Feb-2018

1. M. Kalimuthu, P. Vaishnavi and M. Kishore, "Crop Prediction using Machine Learning," 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), 2020, pp. 926-932,
2. S. S. Kale and P. S. Patil, "A Machine Learning Approach to Predict Crop Yield and Success Rate," 2019 IEEE Pune Section International Conference (PuneCon), 2019.
3. Kavita Jhajhariaa , Pratistha Mathura , Sanchit Jaina, Sukriti Nijhawana(2022).

Crop Yield Prediction using Machine Learning and Deep Learning Techniques. Science Direct, Procedia Computer Science 218 (2023)

406–41.