PREDICTION OF PESTICIDE GOODNESS IN POTATO CROP ACROSS VARIOUS AGRICULTURE LANDS USING MACHINE LEARNING ALGORITHM

Pragatheeswari E[1]
Department of Computer Science and Engineering
Kongu Engineering college
Perundurai, India
epragatheeswari@gmail.com

Saranya S S[2]
Department of Computer Science and Engineering
Kongu Engineering college
Perundurai, India
saranya.sowndarajan@gmail.com

Kavitha M N[3]
Department of Computer Science and Engineering
Kongu Engineering college
Perundurai, India
kavithafeb1@gmail.com

Gopihaa S[4]
Department of Computer Technology - UG
Kongu Engineering college
Perundurai, India
gopihaas.21bcr@kongu.edu

Deepika S[5]
Department of Computer Technology - UG
Kongu Engineering college
Perundurai, India
deepikas.21bcr@kongu.edu

Keerthana M[6]
Department of Computer Technology - UG
Kongu Engineering college
Perundurai, India
keerthanam.21bcr@kongu.edu

***Abstract*— Agriculture is the science of cultivating the land, growing crops and breeding animals. The production of medicinal plants and plant extracts from agriculture also plays a crucial role in the pharmaceutical industries for the development of medications through natural resources. Furthermore, Agriculture provides most of the fabrics such as cotton, wool and leather. India is one of the largest agricultural producers and that is primarily based on its food production and security. In agriculture, farmers use pesticides for protecting the crop from weeds, pests and diseases which leads to better food production. Pesticides are the composition of materials such as plant hormones that regulate the plant growth, defoliants that prevent the plants from dying due to certain factors and also other substances that improve the yield of food products. One of the difficulties of harnessing pesticides is that some kinds of pesticides have an impact on crop productivity. This can be witnessed by predicting the effect of pesticides in crops. The main goal of this paper is to predict the pesticide goodness of potatoes by predicting their yield with the use of pesticide data on various agricultural land utilizing different machine learning algorithms. Machine learning is a computer program which makes the machine learn and train itself by gaining data through experiencing, performing, and understanding the ethics and values of a living ecosystem. The pesticides goodness in crops can be predicted by implementing Random Forest regression, K-Nearest Neighbors regressor, Support Vector Machine regressor and Linear regression algorithms in the 'Indian Crop Production' dataset. This dataset includes crop data along with the year and the amount of pesticides used.**

Keywords — Machine Learning, Prediction, K-Nearest Neighbors, Random Forest, Support Vector Machine, Linear Regressor.

I. INTRODUCTION

Agriculture field is one of the most essential and fastest growing industries in the Indian economy. Every living organism on earth is based on agriculture for their survival. As technology grows, the majority of conventional farming practices have been replaced by the implementation of new tools and techniques for the development of agriculture. Agriculture revolves around five components: cultivable land, machinery, fertilizers and pesticides, irrigation, and high-yielding seed types. Pesticides play a vital role in food production around the world. Pesticides are generally used to exclude or control a variety of household pests that can damage crops, stocks and reduce farm productivity. It is a mixture of substances for preventing, destroying, or lessening the damage caused by the pest. And also, pesticides destroy pests, plant pathogens, and weeds in order to husbandry food without any loss of production.

There are many types of pesticide used in farming such as bio pesticides, synthetic pesticides and chemical pesticides. Chemical pesticides include herbicides, insecticides, fungicides, rodenticides, larvicides, molluscicides, bactericides, and algaecides. Bio-pesticides include microbial pesticides, biochemical pesticides, botanical pesticides, and mineral pesticides. From the recent research, it has been found that the use of bio pesticides leads to the decrease of crop yield. In this paper, we will predict the goodness of auxin herbicides in potatoes, which help in the plant growth. The pesticide data used for this prediction includes synthetic auxin herbicides such as 2,4-D, MCPA and Dicamba. This will stimulate the naturally occurring auxins in the plant cells.

Crop cultivation is under constant threat from various factors such as insect infestations and crop diseases, which negatively impact overall crop productivity. To solve this problem by proposing a solution that focuses on the use of appropriate pesticides and their significant contribution to boosting crop yields.

A. Pesticides in potato crop

The production of potatoes will require a large quantity of pesticides for preventing plant diseases from the growth of microorganisms, bacteria, viruses and so on. Auxin herbicide is a type of herbicide that will tie the plant hormones in the cells of the plant and help the plant growth

within the plants. Auxin herbicides help in the process of tuberization in potatoes. These pesticides include a substance called auxin that acts as a plant hormone in the regulation of plant growth.

These auxins will help the growth of the tuber in the underground stem of the potato. Tuber is a block of storage shaped by the inflammation of the plant stem. It stores the nutrients for the plant to withstand in the extreme weather conditions such as winter, summer and other dry seasons.

To increase crop cultivation, knowledge about pesticide goodness is important to predict the goodness of the pesticide in the crop yield using machine learning algorithms.

B. Machine Learning

Machine Learning is a core element of Artificial Intelligence that includes how machines can break down data, identify patterns and make opinions with low to no mortal intervention. It is a subset of Artificial Intelligence. It is the study of making machines more mortal-like in their actions and opinions by giving them the capability to learn and develop their own programs. This is done with minimal mortal intervention, i.e., no specific programming. The learning process is automated and enhanced grounded on the experiences of the machines throughout the process.

Types of Machine Learning

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

For predicting the goodness of pesticides in potatoes using supervised learning approach. Supervised learning, is a category of machine learning. It's outlined by its usage of marked datasets to train algorithms that to separate data or predict issues directly. There are two types of supervised learning,

- Classification
- Regression

The pesticide goodness can be analysed by predicting the crop yield using regression techniques. Classification is a technique used to predict the categorical data based on the features. In this technique, the model first learns the class labels of the input data and then identifies the category of the test data. Some of the classification algorithms are K-Nearest Neighbor and Support Vector Machine algorithm. These algorithms also work well for regression analysis. Regression is a technique used to predict the numerical data. Some of the regression algorithms are Linear Regression and Random Forest Regression. The crop yield can be predicted by implementing KNN, SVM, linear regression and random forest regression.

II. LITERATURE REVIEW

Naheed Ejaz et.al [1], uses ML/DL algorithm to analyze data about wheat production and also predict how much wheat will can grow. From this study, various ML/DL techniques were applied: one is Neural Network and another one is a mix of Neural Network, Support Vector Machine, and Regression. They used these techniques along with information about the land, crop yield, weather conditions, and farming practices to make predictions about wheat production. This research, helps farmers know how much wheat they can grow.

Dr. Martin Kuradusenge et.al [2], the crop production in agriculture is worsened by the bad climatic changes and other natural calamities. This leads to the issue of food insecurity. The model tries to alleviate those certain factors by the earlier prediction of future crop yield with the help of weather and yield data of Irish potato and maize in the Musanze district in Rwanda. The predictions were done by the implementation of Random forest regression, Polynomial regression and Support Vector Regression algorithms.

Mamunur Rashid et.al [3], predicting the yield of palm oil using machine learning algorithms. First, they give us the lowdown on the current state of palm oil production all over the world. Then, they talk about the info and techniques most commonly used to make these predictions. They also dig deep into the pros and cons of using machine learning to predict crop yields and highlight the challenges it brings to the farming industry. They also explore the future of predicting palm oil yields. This includes using technology like remote sensing to look at how plants are growing and if they're healthy, mapping out fields, and even figuring out the best techniques and tools to use. To sum it up, they present a plan for a clever computer model to predict palm oil yields. This will be super efficient and tackle some of the tricky issues in crop yield prediction.

G Pradeep et.al [4] proposed the crop production forecasting method to address the difficulties such as food shortages, economic instability, inefficient resource allocation, environmental impact, and lower farmer profitability using the proposed machine learning algorithm. The machine learning approach Gradient Boosting Agricultural Yield Prediction which utilize decision trees and gradient descent optimization is used to predict accurate yield predictions about crops with good precision and recall (Accuracy rate - 87.2%, precision - 0.84, recall - 0.90, and F1-Score - 0.87).

Yeshanbele Alebele et.al [5], uses satellite data to predict how much rice a field will produce. There are two types of data were applied: one from satellites that look at the colors of plants, and another from radar satellites. The predictions, done by applying a math technique called Gaussian kernel regression, have been used. They tried different combinations of these data types and found that the best predictions came when they used both the plant color data and the radar data together. This method was better at predicting rice yields compared to other methods. This study shows that using a combination of these two types of satellite data and the Gaussian kernel regression technique can be useful for figuring out how much rice a field will produce.

Ersin Elbasi et.al [6], the paper mainly focused on planting, watering, and harvesting crops. This paper includes a discussion on machine learning in agriculture and impact of changing labels on the accuracy of data analysis algorithms. By analyzing wide-ranging data collected from farms, farmers can make more informed decisions about the affect crop growth. The results show classification accuracy of 99.59% using the Bayes Net algorithm and 99.46% using Naïve Bayes Classifier and Hoeffding tree algorithms. This result will increase the production rate and reduce the cost

for the farms. This study can help future farmers to detect disease early, increase crop production, and reduce prices and food shortages.

Nikita S.Sapike et.al [7], an important issue for the purposes of agricultural planning is yield estimate for the many crops involved in the planning. This paper compares the predictive accuracy for crop yield prediction in ML algorithms. People practice agriculture for years but the results are not effective because of the effect of the crop yield. The idea is to use the eclat technique of association rule mining to create rules and to use genetic algorithms to further refine those rules. The algorithms used are Apriori algorithm, classification, Association Rule Mining technique, Machine Learning.

Umashankar Dewangan et.al [8], Machine learning is a flexible device for farming products because it can forecast crop fields based on factors like position, rainfall, season. In order to prize and synthesize the ways and features are conducted the Systematic Literature Review (SLR) for this work. In our conformance, Artificial Neural Network (ANN) and Random Forest is the most popularized algorithm, which gives the smart performance in closeness and Root Mean Square Error (RMSE). Subordinate model performance in reality, is one of the biggest obstacles. Absence of a large training dataset, increases in overfitting.

Vaibhavi Vanarase et.al [9], the economic system of India is dependent on framing production, the concern food production is essential. It is needed to have good soil quality to get the crop to maximize the crop production. This research paper has found a machine learning algorithm with a system that recommends the best crop based on the nitrogen, phosphorus, and potassium content of the soil. The algorithms are SVM, Naïve bays, and decision tree. Compared to all algorithms, the decision tree gives higher accuracy. The system provides a high profit margin to farmers.

Thomas van Klompenburg et.al [10] extracts and synthesize the algorithms and features in crop yield prediction is performed by a Systematic Literature Review (SLR). Retrieved 567 relevant studies from six electronic databases, from this they have selected 50 studies for further analysis towards its methods and features that they have used. The algorithm used in their studies is Artificial Neural Network ANN). According to their additional studies, Convolutional Neural Networks (CNN), Long-Short Term Memory (LSTM) and Deep Neural Networks (DNN) algorithms from Deep Learning.

R. P. L. Durgabai et.al [11], Farmers have different interests to select suitable Fruit and Vegetable crops. The crop production has reduced due to various factors like pest attack, diseases and climatic conditions. Crop protection is a practice of managing pests, plant diseases, and other damages in agriculture crops. Machine learning is quite effective in the agriculture sector. The crop production has reduced due to various factors like pest attack, diseases and climatic conditions. The different machine learning languages are Pest Management, Machine Learning Algorithms & Agriculture.fter the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

III. RESEARCH METHODOLOGY

This research is focused on predicting the crop yield for potatoes. We have selected a crop named potato in that we us Auxin Herbicide types to predict the yield. In this research, generally uses four types of methodologies. First one is K-Nearest Neighbour Regression, second algorithm is Support Vector Machine, third algorithm is Random Forest Regression and the fourth algorithm is Linear Regression. As soon as you give the input select any method among the KNN, SVM, Random forest Regression, Linear Regression to get the result of yield. The accuracy for the four different algorithms are calculated. By implementing methodologies in dataset we get Accuracy.

A. K-Nearest Neighbor Regression Algorithm:

The K-Nearest Neighbor algorithm is a kind of supervised learning algorithm that is constantly applied in Machine Learning. It is categorized as a non-parametric statistical technique. Non-parametric techniques don't rely on any statistical distribution within the data or make any assumptions about its structure. It uses a non-constant set of parameters.

The distance is determined by utilizing the Euclidean distance equation. The Euclidean distance is the length between the two points. This distance very well may be depicted in any dimension.

$$\text{Euclidean Distance} = \sqrt{(x2 - x1)^2 + (y2 - y1)^2}$$

For all test data points, compute the Euclidean distance of the test data point from distinct training data points.

Find the nearest ‘k’ training data points and find the mean value for the values with the nearest distances.

If (k is 1)

Then assign the single closest value to the output

Else

Continue with Step 1 and 2 until the output is predicted

B. Support Vector Machine Algorithm

Support Vector Machine is one of the most popularized supervised learning algorithms, which is applied for classification as well as regression exceptions. The aim of the SVM algorithm is to deliver the finest line or decision bound that can separate n- dimensional place into classes so that the final data point will be placed in the exact order in the future. This finest decision bound is called a hyperplane. SVM can be applied to all types of supervised learning problems. A SVM regressor is applied for regression analysis. It's especially applied when dealing withnon-linear connections between the input features and the target variable. It gives us the flexibility to outline how major errors respond in our model and will detect a hyperplane in perfected dimensions to fit the data. The error term is rather handled in the constraints, where we set the absolute error

lower than or equal to a specified border, called the maximum error, ϵ(epsilon). The SVR performs a linear regression in the high range of the point space using epsilon ε and tries to reduce the difficulty of the model. The SVR uses a symmetrical loss function, in which both high and low overestimates are corrected in the same way. The main purpose of SVM regression is to discover a hyperplane that dashingly fits the data while minimizing border errors. The SVM regressor predicts the continuous ordered variables, whereas the SVM classifier predicts the categorical variables.

C. Random Forest Algorithm

The random forest algorithm is an ensemble learning approach that combines different machine learning models for prediction. This regression methodology can also be applied for both classification and regression issues. It combines multiple decision trees and produces a final production through average predictions. A decision tree is anon-parametric methodology applied for both classification and regression. It consists of branches, root node and child nodes. The root node indicates the point for the split, the branches indicate the rules and the leaf node indicates the production of the model. In a random forest, the mean of the predictions generated by each individual tree constitutes the final predicting production. By applying an ensemble technique, overfitting is minimized and the model's capacity for conception is enhanced. Assume there are ' n ' variables in the dataset and choose a subset of ' k ' features aimlessly from the n features. A random selection of sample records should be taken. For determining the total number of nodes, apply the optimal splitting approach on 'k' attributes. Continue dividing the nodes into child nodes until the tree reaches its full capability for growth. To train a farther decision tree, choose another set of input data and continue way 1 through 3. Continue repeating this until the'n' decision trees are constructed and trained. Final predictions are made predicated on the normal of the production from the'n' trees.

D. Linear regression Algorithm

A simple linear regression has one independent factor and one dependent factor. The slope and intercept of the best-fit line, which denote the connection between the factors, are estimated by the model. The line that stylish matches the given scatter plot is called the best-fit line. The simple linear regression equation is given by,

$$y = a\ bx$$

where b is the slope of the straight line, x is the independent variable or predictor variable, and y is the intercept or dependent variable. In regression, the diversions between the actual and anticipated values of the production variable are known as residuals. The residuals or random errors, can be calculated by

$$\text{Residuals} = (Y\ \text{prognosticated} - Yi)$$

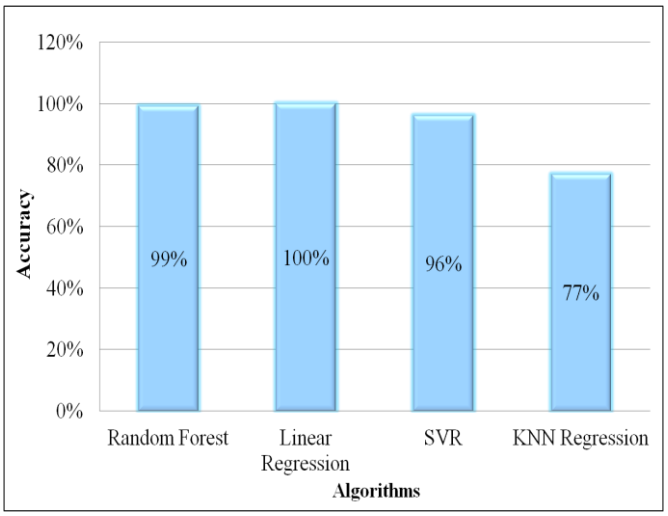
The slope illustrates the variation in the production variable for every single change in the predictor variable, whereas the point of intercept shows the production variable's anticipated value when the predictor variable is zero. The slope of a straight line reflects how important the line on a graph moves vertically over horizontally.

IV. RESULT AND DISCUSSION

The machine learning models were applied throughout the system's entire design. Different features in the dataset have been used for predicting the yield of crop across varied agricultural lands which includes the pesticide details used for cultivation. We've developed a system that predicts the agrarian yield in order to assess the efficacity of pesticides.

The system developed is entirely predicated on the use of pesticides. The variety of pesticides applied for predicting the goodness of pesticide belongs to the family of Auxin herbicides. It includes three types similar as 2,4-Dichlorophenoxyacetic acid, 2- methyl-4-chlorophenoxyacetic acid(MCPA) and 3,6- dichloro-2-methoxybenzoic acid(Dicamba). The Indian Crop Production dataset for predicting the pesticide goodness was taken from Kaggle which is a data repository which contains distinct field datasets.

The data is predicated on several planter cooperatives in the exploration region. The data were gathered from the agricultural eras 1997 to 2020. The machine learning approaches executed for this system includes Random Forest regression, Linear regression, Support vector regressor and K- Nearest neighbor regressor. The predicted results has shown that Random forest and Linear regression models has given the best results with the accuracy of 99 and 100.



V. CONCLUSION

The prediction of Auxin herbicide effectiveness in crop yield enhancement is a crucial aspect of modern agriculture with significant implications for food security and sustainable agricultural methods. Farmers and agricultural specialists can acquire useful insights into the possible influence of various pesticides on crop yield by using modern technology and predictive modelling. The ability to estimate pesticide goodness in terms of crop yield

prediction enables farmers to make informed decisions regarding the selection and application of pesticides.

The application of machine learning methods used for crop yield prediction such as Random forest regression, Linear regression, SVM regression, and K-Nearest Neighbors regression have showed significant possibilities in predicting agricultural productivity. Based on the performance of these models, it is concluded that Linear regression and Random forest regression gives the best results for predicting the yield of potato crop.

These machine learning algorithms not only help to optimize pesticide use. The accuracy of these predictions enables targeted and effective pest control measures, reducing financial losses for farmers.

The observed correlation of greater crop production and increased Auxin herbicide usage illustrates the possible positive influence of conservative herbicide application on agricultural output. The findings imply that, within specific limitations, a strategic and controlled increase in Auxin herbicide implementation. can lead to increased yields.

The efficient utilization of machine learning algorithms in forecasting pesticide goodness for crop yield is a big step forward toward profitable and data-driven agriculture. In conclusion, a larger amount of Auxin herbicide usage may contribute to increased yield. Accurate pesticide goodness prediction using machine learning models, particularly with Auxin herbicides, represents a big step toward precision agriculture.

VI. REFERENCE

[1] Naheed Ejaz and Shabbir Abbasi, "Wheat Yield Prediction Using Neural Network and Integrated SVM-NN with Regression", Pakistan Journal of Engineering Technology and Science (PJETS), vol 8, no.2, December,2018.

[2] Richa Kumari Karn and A. Suresh, "Prediction of Crops Based on a Machine Learning Algorithm", 2023 International Conference on Computer Communication and Informatics (ICCCI), 24 May 2023.

[3] Martin Kurasusenge, Eric Hitimana, Damien Hanyurwimfura, Placide Rukundo, Kambombo Mtonga, Angelique Mukasine, Claudette Uwitonze, Jackson Ngabonziza and Angelique Uwamahoro, "Crop Yield Prediction Using Machine Learning Models: Case of Irish Potato and Maize", School of ICT, College of Science and Technology, University of Rwanda, KN 67, Kigali P.O. Box 3900, Rwanda, vol 13, no.1, 16 January 2023.

[4] Mamunur Rashid, Bifta Sama Bari, Yusri Yusup, Mohamad Anuar Kamaruddin and Nuzhat Khan, "A Comprehensive Review of Crop Yield Prediction Using Machine Learning Approaches With Special Emphasis on Palm Oil Yield Prediction", IEEE Access, vol 9, 22 April 2021.

[5] Anurag Satpathi, Parul Setiya, Bappa Das, Ajeet Sign Nain, Prakash Kumar Jha, Surendra Singh and Shikha Singh, "Comparative Analysis of Statistical and Machine Learning Techniques for Rice Yield Forecasting for Chhattisgarh", India, Department of Agrometeorology, College of Agriculture, G.B. Pant University of Agriculture and Technology, Pantnagar 263153, India, vol 15, no.3, 3 February 2023.

[6] G Pradeep, T Dureen V Rayen, A. Pushpalatha and P. Kavitha Rani, "Effective Crop Yield Prediction Using Gradient Boosting To Improve Agricultural Outcomes", International Conference on Networking and Communications (ICNWC), 25 May 2023.

[7] Yeshanbele Alebele, Wenhui Wang, Weiguo Yu, Xue Zhang, Xia Yao, Yongchao Tian, Yan Zhu, Weixing Cao and Tao Cheng, "Estimation of Crop Yield From Combined Optical and SAR Imagery Using Gaussian Kernel Regression", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol 14, 8 October 2021.

[8] Mummaleti Keerthana, K J M Meghana, Siginamsetty Pravallika and Modepalli Kavitha, "An Ensemble Algorithm for Crop Yield Prediction", 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), 31 March 2021.

[9] Ersin Elbasi, Chamseddine Zaki, Ahmet E. Topcu, Wiem Abdelbaki, Aymen I. Zreikat, Elda Cina, Ahmed Shdefat and Louai Saker, "Crop Prediction Model Using Machine Learning Algorithms", College of Engineering and Technology, American University of the Middle East, Egaila 54200, Kuwait, 16 August 2023.

[10] Nikita S.Sapike and Prof.S.S.Sambare, "Crop yield Prediction Using Machine Learning Algorithm", Journal of Emerging Technologies and Innovative Research (JETIR),vol 7, August 2020.

[11] Umashankar Dewangan, R H Talwekar and Swagota Bera, "Systematic Literature Review on Crop Yield Prediction using Machine & Deep Learning Algorithm", 2022 5th International Conference on Advances in Science and Technology (ICAST), 13 February 2023.

[12] A.P.S Manideep and Dr. Seema Kharb, "A Comparative Analysis of Machine Learning Prediction Techniques for Crop Yield Prediction in India",Turkish Journal of Computer and Mathematics Education, vol 13, no.02, 2022.

[13] Alkha Mohan, "Machine Learning Based Crop Yield Prediction Using Spectral Images", National Institute of Technology Karnataka, Surathkal, January 2022.

[14] R.P.L. Durgabai, P. Bhargavi, Jyothi S, "Pest Management Using Machine Learning Algorithms: A Review", International Journal of Computer Science Engineering and Information Technology Research (IJCSEITR), vol. 8, February 2018.

[15] Indu, Anurag Singh Baghel, Arpit Bhardwaj and Wubshet Ibrahim, "Optimization of Pesticides Spray on Crops in Agriculture using Machine Learning", Department of Computer Science and Engineering, USICT, Gautam Buddha University, Greater Noida, India, vol 2022, 05 September 2022.