CROP YIELD FORECAST MACHINE LEARNING APP

Anusha[1], Sindhu P [2], Supraja A V [3]

[1] Student, Information science and engineering, RajaRajeswari college of Engineering,Bengaluru ,Karnataka, India
[2] Student, Information science and engineering, RajaRajeswari college of Engineering,Bengaluru, Karnataka, India
[3] Student, Information science and engineering, RajaRajeswari college of Engineering,Bengaluru, Karnataka, India

ABSTRACT

The quantity of harvest produced per acre of land is known as crop yield. Understanding this number is essential since it helps evaluate food security and explains why potato prices change annually. the subsequent year. In this project, We'll research several techniques for predicting agricultural yield.This work demonstrates that they may also be completed using machine learning methods like the Random Forest and Kalman Filter methods. To complete the greatest outcomes in crop production prediction, we look at multiple iterations of machine learning algorithms and data processing methodologies in this work. To get an extremely precise yield forecasting percentage, one could use the Random Forest Algorithm with Kalman Filter. Forecasting crop yields is necessary for farmers to make it feasible and to make well-informed decisions on crop management, harvest, and sales. Predicting crop yield with machine learning methods has shown a lot of potential. This study aims to assess the agricultural production predicting performance of a range of methods for machine learning employing historical crop, weather, and soil data. Among Artificial Neural Networks: Artificial Decision Support Systems trees, random forestsas well as support vector machines .The research uses data from several crops, such as wheat, corn, and soybeans, and evaluates the usefulness of metrics such as root mean square error and mean absolute error by all algorithms.

Keyword : - *Crop yield prediction, Random factor, , Kalman filter*

1. INTRODUCTION

To predict agricultural output When utilizing machine learning, one needs to use algorithms in arrange to forecast a given harvest in a specific area. Yield expectations are crucial to farmers because they direct their harvest creation, help them choose which harvests to plant, and help them distribute resources as needed. Agribusinesses, government organizations, and other groups involved in farming might also find it useful. Leveraging machine learning to carry out a harvest crop production prediction involves several steps. Among these methods are:

Data collection: Gathering crucial information that is going to be required to teach The model of machine learning is the first phase. This material may include details regarding the region, kind of soil, types of weather, and along with additional elements that might have an impact agricultural productivity.

Preprocessing of the data: After it has been gathered, the data needs to be cleaned and made suitable for use within a machine learning model. This could entail coding distinct components, scaling the data to a comparable reach, and removing any missing or inaccurate information.

Model selection: Next, a paradigm for machine learning, which is trained to forecast crop yield utilizing the data at hand has to be selected. This could entail comparing the display of different models via cross-approval and selecting the model that functions the finest.

Model training: The preprocessed data should be employed to instruct the chosen model. This means that in order for the role model to accurately predict crop production given the existing data, its boundaries must be adjusted through calculation.

Evaluation of the model: Following training, the model must be assessed to determine how well it functions with sparse information. This might entail testing the model's accuracy with a test set of data that wasn't employed in the holding phase.

Model deployment: To forecast crop output in the real world, a model that performs well in the evaluation stage can be put into use. This could entail integrating the model into an application or using it to provide recommendations to farmers.

In general, Crop yield forecast derived from learning can be a helpful tool for improving harvest creation and helping farmers make wise decisions on their product.

2. LITERATURE SURVEY

[1]2019; S. Bhanumathi, M. Vineeth, and N. Rohit, Crop Yield Forecasting and Effective Fertilizer Use.
This essay shows how, as an agricultural nation, India's economy primarily depends on the growth of agricultural produce and agroindustry products. In agriculture, yield forecasting is a crucial problem. The Random Forest algorithm and the Backpropagation technique are both used to handle the yield creation examination. Both of the harvest creation models were examined in regard of predicting the outcome and using various error rate bounds. The random forest algorithm outperforms the backpropagation technique, founded on the findings.

[2]Crop Utilizing Using Machine Learning to Predict Yield Algorithms, Anakha Venugopal, Aparna S, Jinsu Mani, Rima Mathew, Prof. Vinu Williams, 2021
This essay explains that one of the key factors that contributes most to financial success is agribusiness. AI (ML) may be a key component in discovering a real-world solution that can be applied regarding the problem of crop yield. In arrange to achieve precise predictions and address the inconsistent patterns in temperature and precipitation, various artificial intelligence classifiers, such as Random Forest, Naive Bayes, and logistic regression, are utilized as models.

[3]Machine Acquiring knowledge to predict a crop yields in agriculture, Dr. Mohd. Tahseenul Hasan, Firdous Hina, 2022 This study emphasizes that in order for farmers to increase their yields, they need to have access to modern innovations. Harvest projections that are correct can be conveniently given to acclimatists. an assortment of AI tools were employed to examine the farming-related issues. Because of the dataset's accessibility and the analysis's goal, highlights are still ambiguous Research has indicated that models with additional features could not necessarily result in the optimal realization of return expectations. The models with different quantities of highlights must be compared in arrange to ascertain which works the best. Several computations have been employed in a number of research projects. The results show that some AI .Even though it is hard to definitively say which model is the best, some are actually employed more frequently than others. We were able to determine an anomalous 95% precision by using the combined model.

[4]Mayank Champaneri, Mansing Rathod, Darpan Chachpara, Chaitanya Chandvidkar, Crop Yield Forecast By means of machine learning

The core of the Indian economy is agriculture. The yield of agriculture in India is primarily dependent on atmospheric conditions. Precipitation is mostly necessary for rice cultivation. The best advice is to forecast future yield efficiency and conduct research to help farmers increase the amount of harvests they produce. One of the main issues facing agriculture is yield expectation.

Priya, Muthaiah, and Balamurugan (2018) used a machine learning algorithm to predict the crop's yield. The substantial contribution that agriculture contributes to the growth of the national economy is highlighted in this presentation. The environment and other ecological changes have made farming a much riskier business. AI(ML) is a key technique for achieving effective and useful solutions for this problem. Data mining is also useful for forecasting the formation of harvest yields. Information mining, in its broadest sense, describes the procedure of looking at facts from several angles and condensing it into meaningful insights. The Results show how the Random Forest method may be utilized to obtain a precise harvest yield expectation.

3. PROBLEM STATEMENT

Farmers need to know exactly what harvest yield to expect since it helps them plan their yield creation, decide which yields to grow, and allocate resources effectively. However, projecting agricultural output can be difficult due to the large range of elements that influence it, including soil type, weather patterns, water system practices, and nuisance infestations. Inaccurate assumptions of harvest output might lead to over- or under-creation, which can have detrimental impacts for the environment and economicsThis work's objective is to create an AI model that can estimate agricultural productivity in a particular place with accuracy based on a range of criteria. Not only may the impacts of variables such as soil type be simulated, but atmospheric conditions, and irrigation system practices on crop production, the model should be able to set expectations for different harvests and locations.By accurately predicting crop output, we hope to reduce the risks associated with crop development, simplify irrigation systems for farmers, and help ranchers make knowledgeable decisions about their harvests.

4. PROPOSED WORK

Some of the potential disadvantages of current frameworks may be addressed by a proposed AI-based crop production forecasting system. Potential elements of a suggested framework might be as follows: Extensive data sources:The system may use a broad variety of data sources, including satellite imaging, meteorological data, and other information sources, to increase the forecasts' accuracy.
Simple to use interface: Regardless of the level of experience or expertise in machine learning, ranchers and other partners may easily operate the system thanks to its user-friendly interface.
Economical: The system may be designed to be economically viable, either through the use of open-source tools or by providing membership-based access to the framework at a reasonable cost.High accuracy: The system may be intended to supply high levels of precision in agricultural yield prediction, especially in areas or for crops where data is limited or when the model needs to describe intricate interactions between several variables.

5. IMPLIMENTATION

5.1 Methodology

The processes that an agricultural yield prediction system based on machine learning would normally take include data collection, preprocessing, model selection, training, evaluation, and deployment. **Data collection**:

The most crucial step in implementing an AI-based framework for harvest yield forecasting is gathering relevant data for the model. Information regarding the region, the kind of soil, the climate, irrigation methods, plus additional components that could affect crop productivity may be included in this data.

Preprocessing data:

After It's been collected, the data needs to be cleaned and made decent enough to be utilized within an illustration of artificial intelligence.This22944could entail coding all variables, scaling www.ijariie.cothe information to a comparable reach, and removing any4 incorrect or missing information.

Model selection:

Next, a model of AI that can forecast crop yield ought to be chosen using the data at hand. This could entail selecting the model that performs the best after comparing the display of several models via cross-approval.

Model training:

After a model has been selected, the preprocessed information ought to used to prepare it. To do this, a computation is utilized to modify the model's bounds so that, given the data, it can accurately predict crop yield.

Evaluation of the model:

Following its preparation, the model needs to be evaluated to determine how well it functions with subtle data. This could involve surveying the model's accuracy using a test set of data that wasn't used during the preparation cycle or using it to provide recommendations to ranchers.

Model deployment:

If the model shows promise in the evaluation phase, it is often deployed to a certifiable setting to generate crop yield predictions. This could entail integrating the model with an application.

6. Proposed system architecture

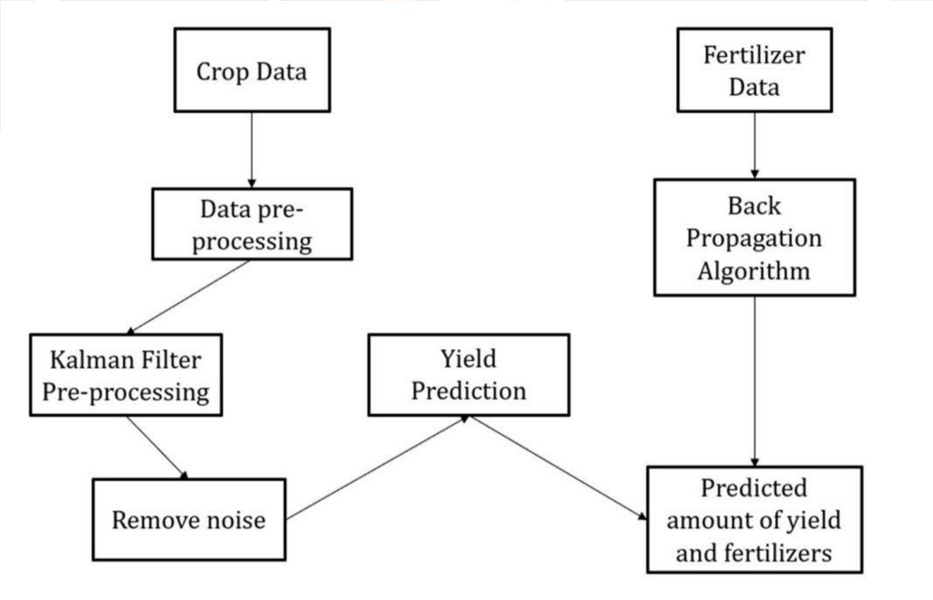

Fig 6.1 Proposed system architecture

7.[22944]APPLY CASE DIAGRAM

The most basic kind of use case diagram is one that depicts how a user interacts with the framework and the connections among them and the many use cases they are involved in. A system's numerous use cases and user types can be distinguished using a use case diagram. It is commonly employed alongside other kinds of diagrams. An illustration of the use case is shown below.

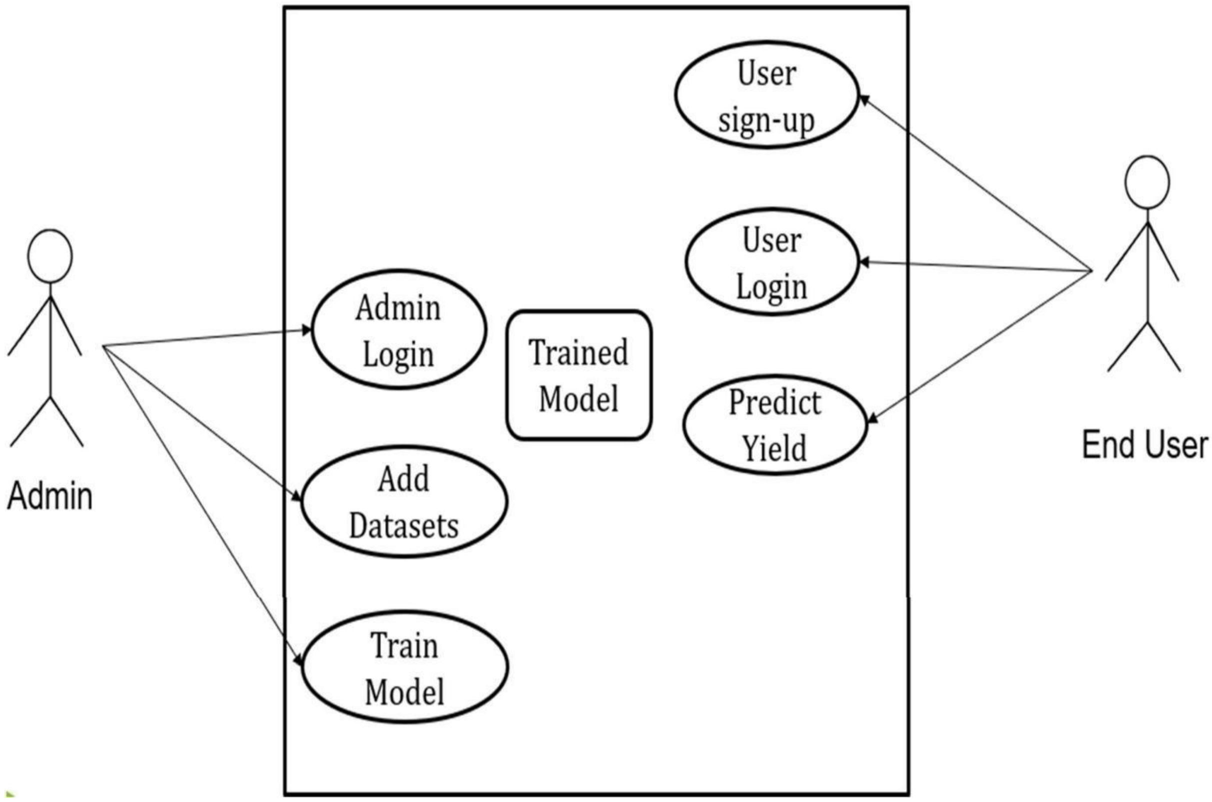

Fig 7.1 Use case diagram of proposed system

8. MODEL EVALUTION

In this module, the trained model's performance is assessed utilizing a variety of measures, including F1 score, recall, accuracy, and precision.

Choosing an evaluation metric to gauge The evaluation module begins with the model's performance.. For regression issues, including crop yield prediction, RMSE, MSE, and RMSE were the root mean square errors. R-squared (R2) are often used evaluation metrics.

Once22944the evaluation measure has been selected,www.ijariie.cothe trained modem 1 is evaluated on an alternative testing 6 dataset that was not used during the model training phase. The real-world circumstances that the purpose of the model is applied to should be reflected in the testing dataset.

Model is evaluated by contrasting the real yields of each crop sample in the testing dataset with the crop yields predicted by the prototype. The discrepancies between the crop yields, both predicted and actual are then utilized to compute the evaluation metric.

An indicator of how effectively the model forecasts crop yield is supplied by the evaluation metric. If the event that the model's performance is deemed unacceptable, either the model or the dataset utilized for training it could be altered, or the evaluation process may be repeated until a suitable model is generated.

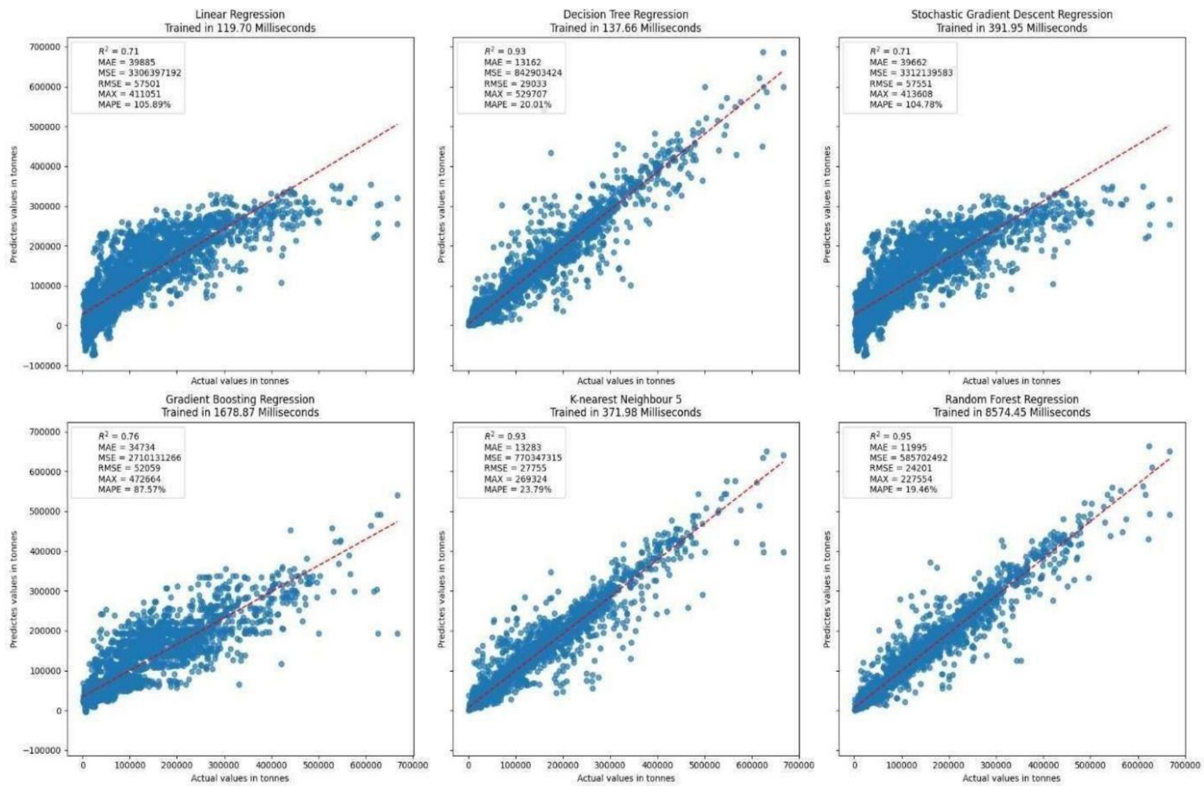

Fig 8.1 Predictes values in tonnes vs actual values in tonnes

9. MODEL SELECTION

In this session, you will choose the machine learning method that most closely matches the provided problem and data. Common algorithms include random forests, decision trees, neural systems, and linear regression.

Crop yield prediction can be solved using a variety of machine learning techniques, such as neural networks, decision trees, random forests, and linear regression. The type of issue, the quantity and intricacy of the data, and the available processing capacity are some of the factors that go into selecting the optimal solution.

Creating training, validation, and testing sets from the available data is a popular method of selecting models. Various models are instructed using the training set using various hyperparameters and algorithms The effectiveness of every model is then assessed using the validation set, and the best model is chosen using a selected evaluation metric (e.g., mean squared error or R-squared). To anticipate the chosen model's performance on fresh, untested data, it is lastly assessed on the testing set.Using cross-validation, which splits the data into numerous folds and trains and assesses the models on various combinations of training and validation sets, is an additional method for choosing a model. This method can lessen the chance of overfitting and offer a more trustworthy approximation of various models' performance.

Model selection may entail choosing the optimal hyperparameters for every algorithm in addition to the optimal algorithm. Hyperparameters are numbers that are chosen before the the model is instructed and have a big influence on how well it works. Hyperparameters that can be adjusted to enhance model performance include the number of hidden layers in a neural network or the regularization strength in linear regression.

10. RESULTS AND DISCUSSION

First, several regression methods are imported from scikit-learn, including K-nearest Neighbor 5, Random Forest, Decision Tree, Gradient Boosting, Stochastic Gradient Descent, and Linear Regression.
The dataset is then split into sets for instruction and testing using the scikit-learn train_test_split function. Next, the algorithm goes through each regression technique, training them on the training set using the fit method. The trained models' predictions are subsequently generated on the testing set by utilizing the predict method.For every regression algorithm, a plot is produced using the plot_regression_results function.
The y-axis of the figure shows the expected values, while the x-axis shows the testing set's actual values. In addition to the regression line, other performance metrics, including R^2, MAE, MSE, RMSE, and MAX

11. CONCLUSION

In conclusion, recent trials using machine learning techniques for crop yield prediction have produced encouraging results. Machine learning algorithms can utilize past crop data, weather patterns, soil qualities, and other pertinent information to properly anticipate agricultural output. This gives farmers the knowledge and ability to make wiser choices when handling their crops. Crop production prediction has been the subject of numerous machine learning approaches, such as random forests, neural networks, as well as a support vector machines,decision trees, and linear regression.Every one of these strategies has advantages and disadvantages, and the best strategy will rely on the particulars of the problem at hand additionally as the data that is currently available.

12. REFERENCES

[1] Vinu Williams, 2021; Anakha Venugopal, Aparna S, Jinsu Mani, Rima Mathew, International Journal of Engineering Research & Technology (IJERT) NCREIS-2021 (Volume 09 – Issue 13) CropYield Prediction using Machine Learning Algorithms

[2] Vineeth, M., Bhanumathi, S., and Rohit, N. (2019). Prognosticating Crop Yield and Optimizing Fertilizer Use. 2019 International Conference on Communication and Signal Processing (ICCSP), 0769–0773.

[3]In the 2022 Third International Conference on Intelligent Engineering and Management (ICIEM), pp. 754757, S. Vashisht, P. Kumar, and M. C. Trivedi present "Improvised Extreme Learning Machine for Crop Yield Prediction." doi: 10.1109/ICIEM54221.2022.9853054.

[4] "Crop Yield Prediction Using Machine Learning Algorithm," 2021 4th International Conference on Computing and 0Communications Technologies (ICCCT), 2021, pp. 611-616, doi:10.1109/ICCCT53315.2021.9711853, R. J., V. K. G. Kalaiselvi, A. Sheela, D. S. D, and J. G.

[5] Agricultural Analysis and Crop Yield Prediction of Habiganj using Multispectral Bands of Satellite Imagery with Machine Learning, 11th International Conference on Electrical and Computer Engineering (ICECE), 2020, pp. 21–24; F. Shahrin, L. Zahin, R. Rahman, A. J. Hossain, A. H. Kaf, and A. K. M. Abdul MalekAzad