CROP YIELD PREDICTION AND REMEDIES RECOMMENDATION USING VARIOUS FEATURE SELECTION TECHNIQUES IN MACHINE LEARNING

Sairaj V[1], S Venkata Sai Tanish[2], VS Venkatesh[3], Rishitha Panyam[4]

[1] *Student, Computer Science and Engineering, BMSIT&M, Karnataka, India*
[2] *Student, Computer Science and Engineering, BMSIT&M, Karnataka, India*
[3] *Student, Computer Science and Engineering, BMSIT&M, Karnataka, India*
[4] *Student, Computer Science and Engineering, BMSIT&M, Karnataka, India*

ABSTRACT

Agriculture is the foundation of many countries' economies, particularly in India. The young generation who are new to farming may confront the challenge of not understanding what to sow and what to reap benefit from. This is a problem that has to be addressed, and it is one that we are addressing. Predicting the proper crop and production will aid in making better decisions, reducing losses and managing the risk of price fluctuations. The existing system is not deployed, unlike ours, which is done by applying classification and regression algorithms to calculate crop type recommendations and yield predictions. Agricultural industries must use machine learning algorithms to anticipate the crop from a given dataset. The supervised machine learning technique is used to analyse a dataset in order to capture information from multiple sources, such as variable identification, uni-variate analysis, bi-variate and multi-variate analysis, missing value treatments, and so on. A comparison of machine learning algorithms was conducted in order to identify which algorithm was more accurate in predicting the best harvest. The results show that the proposed machine learning algorithm technique has the best accuracy when comparing entropy calculation, precision, Recall, F1 Score, Sensitivity, Specificity, and Entropy.

Keyword:- *Machine Learning algorithms, recommendations, predictions*

1. INTRODUCTION

The history of agriculture in India dates back to the Indus Valley Civilization Era. India ranks second in this sector. Agriculture and allied sectors like forestry and fisheries account for 15.4 percent of the GDP (gross domestic product) with about 31 percent of the workforce. India ranks first globally with the highest net cropped area followed by US and China. Agriculture is demographically the broadest economic sector and plays a significant role in the overall socio-economic fabric of India. Due to the revolution in industrialization, the economic contribution of agriculture to India's GDP is steadily declining with the country's broad-based economic growth. The problem that the Indian Agriculture sector is facing is the integration of technology to bring the desired outputs. With the advent of new technologies and overuse of non-renewable energy resources patterns of rainfall and temperature are disturbed. The inconsistent trends developed from the side effects of global warming make it cumbersome for the farmers to clearly predict the temperature and rainfall patterns thus affecting their crop yield productivity. In order to perform accurate prediction and handle inconsistent trends in temperature and rainfall various machine learning algorithms like RNN, LSTM, etc can be applied to get a pattern. It will complement the agricultural growth in India and all together augment the ease of living for farmers. In past, many researchers have applied machine learning techniques to enhance agricultural growth of the country.

This paper focuses on predicting the yield of the crop by applying various machine learning techniques. The outcome of these techniques is compared on the basis of mean absolute error. The prediction made by machine learning algorithms will help the farmers to decide which crop to grow to get the maximum yield by considering factors like temperature, rainfall, area, etc Objectives to be followed in the future are given below:

1. Depending on the dissimilar crop feature divisions, the modulating factor values of ML algorithms differ to attain perfect approximation.
2. When the quantity of input elements is reduced, ANN is utilized. The optimal feature was being empirically selected for appropriate crop yield estimation.
3. The advantage of ML method regression is to avoid difficulties of using a linear function in large output sample space and optimization of complex problems transformed into simple linear function optimization.
4. ML algorithm can be executed with an enormous soil dataset for crop yield estimation.
5. The ML techniques, through observation of the agricultural fields, provided the necessary support to the farmers in increasing crop production to a great extent.

Fig -1: Example figure

2. LITERATURE SURVEY

Aruvansh Nigam et al. (2019) a model which employs ML techniques, including Linear Regression, Decision Tree, and Random Forest, for crop yield prediction[1]. Utilizing satellite imagery and weather data, the researchers extracted features to predict crop yield. Results indicate that Random Forest outperformed other ML algorithms in crop yield prediction. The research suggests that ML-based methods provide a promising approach for accurate and efficient crop yield prediction. The study offers insights into the potential of ML algorithms for optimizing agricultural practices and improving crop yield forecasts. Among the ML algorithms used, the random forest algorithm yielded the highest accuracy, achieving a remarkable accuracy rate of 92%.

Leo Brieman et al.(2001) is a study which elucidates the Random Forest algorithm, its properties, and applications[2]. It discusses how Random Forest reduces overfitting and error rates compared to individual decision trees. The paper details the algorithm's construction, illustrating how it samples the data and builds multiple decision trees. Breiman also explains the process of aggregating the predictions of these trees to achieve the final prediction. The research concludes by demonstrating the effectiveness of Random Forest through empirical examples, showcasing its robustness and efficiency in various domains.

Priya, P. et al. (2020) has various machine learning techniques such as Linear Regression, Decision Tree, and Random Forest are employed in the study to predict crop yield[3]. By integrating satellite imagery and weather data, the researchers extract features crucial for crop yield prediction. The paper evaluates the performance of the machine learning algorithms on the dataset, analyzing their accuracy and efficiency. Results indicate that the Random Forest algorithm outperformed other machine learning algorithms in crop yield prediction. The research suggests that

machine learning-based methods offer a promising approach for accurate and efficient crop yield prediction. The study provides insights into the potential of machine learning algorithms for optimizing agricultural practices and improving crop yield forecasts. It recommends integrating machine learning techniques into agricultural systems to enhance productivity and sustainability.

Shruti Kulkarni et al. (2018) proposed a study which utilizes a neural network model to predict crop yield, emphasizing its potential in agriculture[4]. By integrating historical crop data, weather information, and soil conditions, the researchers extract features crucial for predicting crop yield. The paper evaluates the performance of the neural network model on the dataset, analyzing its accuracy and efficiency. Results indicate that the neural network model demonstrates promising accuracy in crop yield prediction. The research suggests that predictive analysis using a neural network model provides a robust approach for accurate and efficient crop yield prediction. It offers insights into the potential of neural network models for optimizing agricultural practices and improving crop yield forecasts. The study recommends integrating predictive analysis using neural network models into agricultural systems to enhance productivity and sustainability.

K. P. K. Devan et al. (2023) study proposes a novel approach that integrates machine learning algorithms to predict crop yield accurately and recommend suitable fertilizers[5]. Various hybrid machine learning algorithms are employed, combining the strengths of different models. By incorporating historical crop data, weather information, and soil conditions, the system extracts essential features for predicting crop yield. The paper evaluates the performance of the hybrid machine learning algorithms on the dataset, analyzing their accuracy and efficiency. Results show promising accuracy in crop yield prediction and fertilizer recommendation. The research suggests that the proposed system offers an effective approach for precise and efficient crop yield prediction and fertilizer recommendation. It provides valuable insights into the potential of hybrid machine learning algorithms for optimizing agricultural practices, improving crop yield forecasts, and enhancing sustainability. with other cluster approaches.

Bang, S. et al. (2019) study proposes a method that utilizes temperature and rainfall parameters, predicted through ARMA, SARIMA, and ARMAX models, for accurate crop yield prediction[6]. By incorporating these parameters, the system utilizes fuzzy logic to predict crop yield. The paper evaluates the performance of the system on the dataset, analyzing its accuracy and efficiency. Results demonstrate promising accuracy in crop yield prediction. The research suggests that the proposed system provides an effective approach for precise and efficient crop yield prediction. It offers valuable insights into the potential of fuzzy logic-based systems for optimizing agricultural practices, improving crop yield forecasts, and enhancing sustainability.

3. METHODOLOGY

Leo Brieman [2], focuses on accuracy and power and the interaction of the random forest algorithm. The random forest algorithm creates tree trees from different data samples and predicts data from each subset and by voting provides the best system response. Random Forest used a wallet method to train data. In order to improve accuracy, random injections should reduce affinity while maintaining strength. Priya P [3], used the prediction of crop yields using only a random forest divider. Various factors are taken into account such as rainfall, temperature and time of year to predict crop yields. Some machine learning algorithms have not been applied to databases. In the absence of other algorithms, comparisons and measurements were not available and therefore could not provide a valid algorithm. Shruti [4], in theory, has described various machine learning techniques that can be used in a variety of predictive environments. However, their function fails to use any algorithms and thus cannot provide a clear understanding of the functionality of the proposed function.

3.1 Drawbacks

- Some machine learning algorithms have not been applied to databases.
- Comparisons and measurements were not available and therefore could not provide a valid algorithm.
- However, their function fails to use any algorithms and thus cannot provide a clear understanding of the functionality of the proposed function.

As a decision support tool for Crop Yield Prediction (CYP), which includes assisting decisions on which crops to plant and what to do during the growth season of the crops, machine learning (ML) plays a vital part in this study. This study combines a variety of machine learning approaches to forecast the crop's output. Based on mean absolute error, these techniques' results are contrasted. By taking into account variables like temperature, rainfall, area, and

other factors, the predictions made by machine learning algorithms will assist farmers in choosing which crop to grow to get the greatest yield.

3.2 Benefits

- Automatic methods of distinguishing plant diseases also help to take action after detecting the symptoms of leaf diseases.
- By adjusting the model architecture, we transformed it into a classification model, which obtained a higher accuracy than state-of-the- art models.

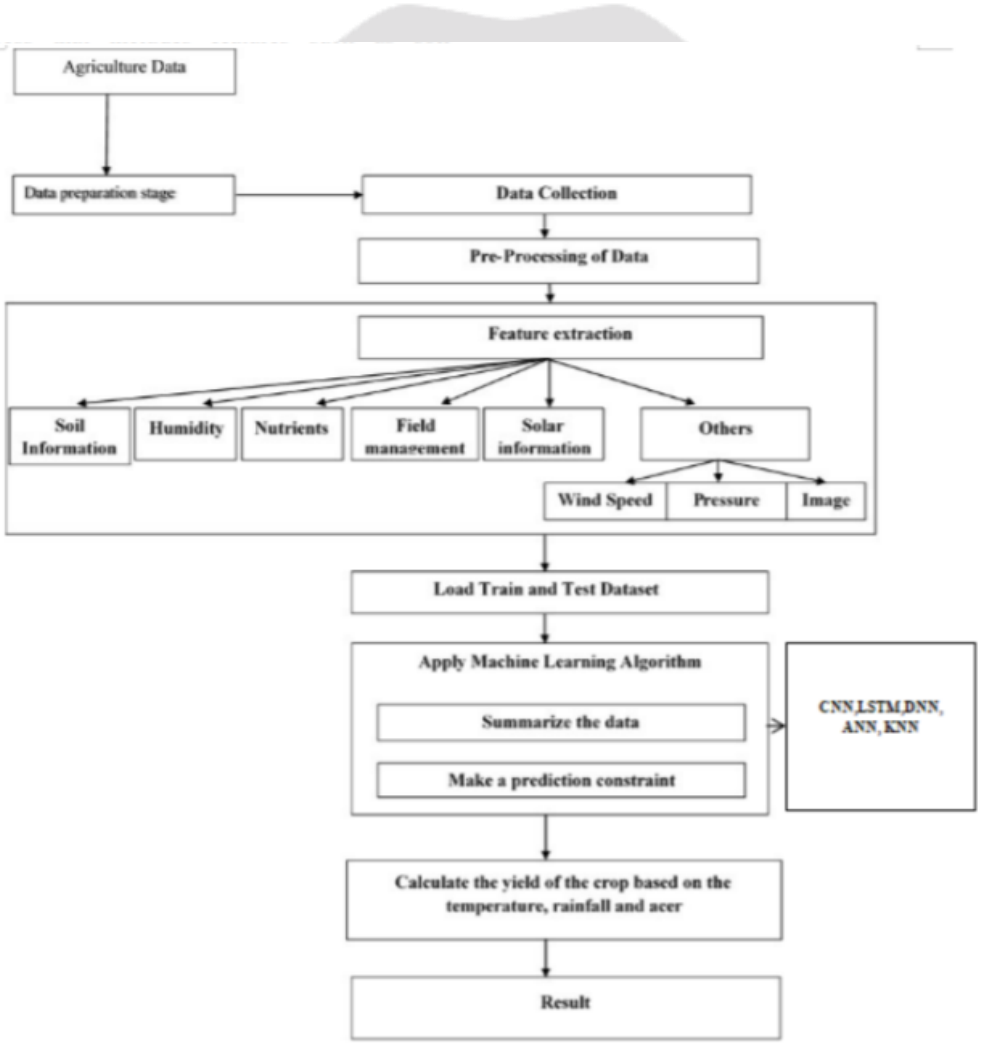

Fig -2: System Architecture

3.3 Modules

- Data exploration: using this module we will load data into system
- Processing: Using the module we will read data for processing

- Model generation: Algorithms accuracy calculated
- User signup & login: Using this module will get registration and login
- User input: Using this module will give input for prediction
- Prediction: final predicted displayed

4. IMPLEMENTATION

- Inception resent V2: Inception-ResNet-v2 is a convolutional neural network that is trained on more than a million images from the ImageNet database [1]. The network is 164 layers deep and can classify images into 1000 object categories, such as keyboard, mouse, pencil, and many animals.
- AlexNet: AlexNet architecture consists of 5 convolutional layers, 3 max-pooling layers, 2 normalization layers, 2 fully connected layers, and 1 softmax layer. 2. Each convolutional layer consists of convolutional filters and a nonlinear activation function ReLU. 3. The pooling layers are used to perform max pooling.
- CNN: A CNN is a kind of network architecture for deep learning algorithms and is specifically used for image recognition and tasks that involve the processing of pixel data. There are other types of neural networks in deep learning, but for identifying and recognizing objects, CNNs are the network architecture of choice.
- Random Forest: Random forest is a commonly-used machine learning algorithm trademarked by Leo Breiman and Adele Cutler, which combines the output of multiple decision trees to reach a single result. Its ease of use and flexibility have fueled its adoption, as it handles both classification and regression problems.
- Decision Tree: A decision tree is a non-parametric supervised learning algorithm, which is utilized for both classification and regression tasks. It has a hierarchical, tree structure, which consists of a root node, branches, internal nodes and leaf nodes.
- KNN: KNN is one of the simplest forms of machine learning algorithms mostly used for classification. It classifies the data point on how its neighbor is classified. KNN classifies the new data points based on the similarity measure of the earlier stored data points.
- Naive Bayes: The Naïve Bayes classifier is a supervised machine learning algorithm, which is used for classification tasks, like text classification. It is also part of a family of generative learning algorithms, meaning that it seeks to model the distribution of inputs of a given class or category.
- Simple RNN: A RNN is designed to mimic the human way of processing sequences: we consider the entire sentence when forming a response instead of words by themselves. For example, consider the following sentence: “The concert was boring for the first 15 minutes while the band warmed up but then was terribly exciting.”
- LSTM – ANN: Long short-term memory (LSTM) is an artificial neural network used in the fields of artificial intelligence and deep learning. Unlike standard feedforward neural networks, LSTM has feedback connections.
- Linear Regressor: Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.
- XGBoost Regressor: XGBoost is an efficient implementation of gradient boosting that can be used for regression predictive modeling. How to evaluate an XGBoost regression model using the best practice

technique of repeated k-fold cross-validation. How to fit a final model and use it to make a prediction on new data.

- Voting classifier: A voting classifier is a machine learning estimator that trains various base models or estimators and predicts on the basis of aggregating the findings of each base estimator. The aggregating criteria can be combined decision of voting for each estimator output.

5. CONCLUSION

The present research work discussed about the variety of features that are mainly dependent on the data availability and each of the research will investigated CYP using ML algorithms that differed from the features. The features were chosen based upon the geological position, scale, and crop features and these choices were mainly dependent upon the data-set availability, but the more features usage was not always giving better results. Therefore, finding the fewer best performing features were tested that also have been utilized for the studies. Most of the exiting models utilized Neural networks, random forests, KNN regression techniques for CYP and a variety of ML techniques were also used for best prediction. From the studies most of the common algorithms used were CNN, LSTM, DNN algorithms but still improvement was still required further in CYP. The present research shows several existing models that consider elements such as temperature, weather condition, performing models for the effective crop yield prediction. Ultimately, the experimental study showed the combination of ML with the agricultural domain field for improving the advancement in crop prediction. However, still more improvement in feature selection was required in terms of temperature variation aspects effects on agriculture. In the further studies, the key possibility that should be concentrated such as firstly the delay to border topographical areas required additional-explicit treatment. Next, a nonparametric portion of the model using machine learning algorithm and thirdly, using features from deterministic crop models to get perfect statistical CO2 fertilization. By following above-mentioned objectives, the crop yield estimation would be improved by further researchers. Additionally, in the crop yield estimation, fertilizer should also be considered for executing soil forecasts that agriculturalist to make a better judgment based on the situation of low crop yield estimation. Based on the outcomes obtained for the study further we need to build and develop a model based on DL for CYP.

6. REFERENCES

[1]. Aruvansh Nigam, Saksham Garg, Archit Agrawal Crop Yield Prediction using ML Algorithms, 2019 Fifth International Conference on Image Information Processing (ICIIP) ,2019

[2].Leo Brieman, Random Forests, Published in: Machine Learning 45 (2001) 1, 5-32, tps://www.scirp.org/(S(czeh2tfqw2orz553k1w0r45))/journal/home.aspx?journal id=30" \t "_blank" Journal of Biomedical Science and Engineering, blank" Vol.9 No.5 , 2001

[3]. Priya, P., Muthaiah, U., Balamurugan, M.Predicting Yield of the Crop Using Machine Learning Algorithm,, Volume 9 Issue 4, April 2020 http://www.ijsr.net2015 www.ijsr.net2015

[4]. Shruti Kulkarni, Shahnawaz Mandal, G Srivatsa Sharma, “Predictive Analysis to Improve Crop Yield using a Neural Network Model” IEEE-2018

[5]. K. P. K. Devan, B. Swetha, P. Uma Sruthi and S. Varshini, "Crop Yield Prediction and Fertilizer Recommendation System Using Hybrid Machine Learning Algorithms," *2023* IEEE 12th International Conference on Communication Systems and Network Technologies (CSNT), Bhopal, India, 2023, pp. 171-175,

[6]. Bang, S., Bishnoi, R., Chauhan, A. S., Dixit, A. K., & Chawla, I. (2019, August). Fuzzy Logic based Crop Yield Prediction using Temperature and Rainfall parameters predicted through ARMA, SARIMA, and ARMAX models. In 2019 Twelfth International Conference on Contemporary Computing (IC3) (pp. 1-6). IEEE.