Journal of
Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.7765868

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

Beesakkagari Jagadeesh Yadav
GITAM University, Hyderabad.
jagadeeshyadavb134@gmail.com

Dr.M.Akkalakshmi
Professor, GITAM University, Hyderabad.
amuddana@gitam.edu

Abstract: The use of machine learning techniques in predicting crop yield has become increasingly important in recent years, as farmers face rapid changes in environmental conditions that make it difficult to make accurate predictions using traditional methods. Efficient feature selection methods are crucial to ensure that machine learning models work accurately, and the selection of appropriate features is key to achieving high levels of precision. Various machine learning techniques can be used to predict crop yield, and the results of these techniques can be compared based on their mean absolute error. By considering factors such as temperature, rainfall, and area, farmers can use the predictions made by these algorithms to decide which crop to grow to achieve the maximum yield. Different machine learning algorithms may have varying modulating factor values, depending on the specific features of the crop being analyzed. Artificial neural networks (ANNs) may be used when the quantity of input elements is reduced, and optimal features can be empirically selected to ensure accurate crop yield estimation. By using machine learning algorithms to analyze large soil datasets, farmers can obtain valuable insights that can help them increase crop production significantly. Ultimately, the use of machine learning in agriculture can support sustainable food production and help to meet the increasing demand for food worldwide.
Keywords: Machine Learning. Feature Selection, Agriculture, classification, crop prediction, Recommendation.

I. INTRODUCTION

1.1 AIM

Machine learning algorithms can provide decision support for crop yield prediction by analyzing various factors such as temperature, rainfall, and soil conditions. This information can be used by farmers to make informed decisions on which crops to grow and how to manage them during the growing season to maximize yield. ML algorithms can also help farmers identify potential problems early on, such as pest infestations or disease outbreaks, and take proactive measures to address them. Machine learning can be a valuable tool in supporting decisions related to crop yield prediction and improving agricultural productivity.

1.2 INTRODUCTION

Agriculture has been a significant sector in India's economy for centuries, and it continues to be a vital source of income for millions of people in the country. However, with the changing patterns of rainfall and temperature due to climate change, farmers are facing many challenges in predicting and managing their crop yield. The integration of technology, including machine learning, can help to address these challenges and improve agricultural productivity. Machine learning algorithms such as recurrent neural networks (RNNs) and long short-term memory (LSTM) models can be applied to historical data to identify patterns and predict future trends in temperature and rainfall. This information can be used by farmers to make informed decisions about crop selection, planting times, and irrigation practices, which can ultimately lead to higher crop yields. Several researchers have already applied machine learning techniques to agriculture in India, with promising results. For example, machine learning algorithms have been used to predict the yield of crops such as wheat and rice, as well as to identify crop diseases and pests. The integration of technology and machine learning has the potential to significantly improve agricultural productivity in India and provide farmers with more accurate and reliable predictions for their crop yields. By leveraging these tools, India can continue to grow and develop its agricultural sector in a sustainable and efficient manner.

1.3 OBJECTIVES:

This paper discusses the use of machine learning techniques to predict crop yield based on various factors such as temperature, rainfall, and area. The main objective is to help farmers determine which crops to grow for maximum yield. In the future, the paper suggests that modulating factor values for ML algorithms should be adjusted based on the crop features being analyzed. Additionally, when input elements are reduced, artificial neural networks (ANNs) should be used for optimal feature selection. The paper also proposes the use of large soil datasets for crop yield estimation and suggests that ML techniques can provide valuable support to farmers in increasing crop production. Ultimately, the goal is to help farmers make informed decisions about which crops to grow in different environmental conditions to maximize yield.

1.3 EXISTING MODELS:

The existing research on crop yield prediction using machine learning has faced several challenges.

- Firstly, the creation, repair, and maintenance of ML algorithms can be costly due to their complexity.
- Secondly, some ML techniques used for crop yield prediction failed to obtain better results statistically because they combined input and output data for multiple crops such as mustard and wheat.
- Thirdly, linear regression models were not effective in predicting crop yield accurately in complex situations involving extreme value data and nonlinear data.
- Fourthly, K-NN models used for classification in yield prediction faced performance issues due to nonlinearity and high adaptability issues. The models were also operated in a locality model, which increased the dimensionality of the input vector and made classification confusing.
- Lastly, insufficient data availability has resulted in inappropriate decision-making during classification for crop yield estimation.

1.3 METHODOLOGY:

This paper focuses on the practical application of machine learning algorithms for crop yield prediction and proposes an approach that considers all relevant features, rather than just one at a time. The study also addresses the challenge of inconsistent data from rainfall and temperature datasets by deriving a consistent trend. Additionally, the paper proposes remedies to improve crop yield prediction. The research shows that common algorithms such as CNN, LSTM, and DNN can be used for crop yield prediction, but further improvements are required in this field. Accurate farming requires real-time data on weather, air quality, soil, crop maturity, machinery, labor costs, and current data availability. Predictive analysis can help make wise decisions in agriculture and overcome the declining crop production in the face of a growing population.

Machine learning is a promising tool in agriculture, as it enables computers to learn how to perform specific tasks without explicit programming. It involves the computerized study of data to solve problems and improve algorithms. Training data can be used to improve algorithms for tasks where there are several possible answers. For instance, the MNIST dataset can be used to train digital character recognition systems. The study proposes an approach that leverages machine learning to improve crop yield prediction and increase the efficiency and productivity of agriculture.

II. RELATED WORK

2.1 Crop Yield Prediction Using Machine Learning Algorithms:

This research paper aims to predict crop yield using machine learning algorithms to improve the income scenario of agriculture in India. The study utilizes various machine learning techniques and compares their outcomes based on mean absolute error. The prediction results can assist farmers in deciding which crop to grow based on factors such as temperature, rainfall, and area, to achieve maximum yield.

2.2 Predicting Yield of the Crop Using Machine Learning Algorithm:

The paper discusses the use of machine learning algorithms, specifically Random Forest, to predict crop yield based on historical weather and soil data. The authors highlight the importance of agriculture in a country's economy and the threat posed by climate and environmental changes. The proposed solution involves using machine learning to provide practical and effective solutions to the problem of predicting crop yield.

The study uses real data from Tamil Nadu and builds models based on the available data. The models are then tested with sample data to evaluate their accuracy. The paper concludes that Random Forest is a powerful and popular algorithm for predicting crop yield accurately. The predicted yield can be used by farmers to make informed decisions before cultivating crops. The study highlights the potential of machine learning algorithms in the agriculture sector. Predicting crop yield accurately can help farmers optimize their crop production and increase their profitability.

2.3 Applications of machine learning techniques in agricultural crop production:

The paper reviews the relevance of machine learning techniques in the domain of agricultural crop production. The authors highlight the importance of accurate and timely forecasts of crop production for policy decisions related to import-export, pricing, marketing, and distribution. However, prior estimates are often subjective and require descriptive assessment based on qualitative factors. Thus, there is a need to develop statistically sound objective predictions of crop production.

The development of computing and information storage has provided a large amount of data, but extracting knowledge from this raw data can be challenging. Therefore, the paper discusses the use of machine learning techniques such as artificial neural networks, Information Fuzzy Network, Decision Tree, Regression Analysis, Bayesian belief network, Time series analysis, Markov chain model, k-means clustering, k nearest neighbor, and support vector machine in the agricultural domain to evaluate the relationship between various variables present in the database. The paper highlights the potential of machine learning techniques in crop production management. These techniques can help provide objective estimates of crop production and assist in making important policy decisions.

2.4 Predictive Analysis to Improve Crop Yield using a Neural Network Model:

The paper discusses the development of a data-driven model using a hybrid neural network to predict crop yield in several districts based on historic soil and rainfall data. The authors highlight the importance of agriculture in feeding the population and contributing to GDP, and the role of various factors such as soil properties, climate, elevation, and irrigation technique in determining crop yield. However, technological developments have fallen short in estimating yield based on the joint dependence of these factors.

The study focuses on rice crop, and the hybrid neural network model is designed to identify optimal combinations of soil parameters and rainfall patterns in a selected region to predict crop yield over seasons. The model uses a Time-Series approach in Supervised Learning for the predictive analysis of rainfall, and a Recurrent Neural Network as a branch of Machine Learning for the final prediction of crop yield. By using two inter-communicating data-driven models working at the backend, the final predictions were successful in depicting the interdependence between soil parameters for yield and weather attributes. The study highlights the potential of data-driven models using neural network techniques in predicting crop yield accurately. These models can help farmers optimize their crop production by making informed decisions about soil parameters and weather attributes to achieve the highest possible yield.

2.5 CRY—an improved crop yield prediction model using bee hive clustering approach for agricultural data sets:

Agricultural researchers over the world insist on the need for an efficient mechanism to predict and The CRY model proposed in this research paper is designed to predict and improve crop growth through an adaptive cluster approach based on historical crop data. The model uses a bee hive modeling approach to analyze and classify crops based on their growth patterns and yield. The classified dataset is then tested using Clementine over existing crop domain knowledge. The performance of CRY is compared to other cluster approaches, and the results show that CRY performs better in predicting crop yield. This model has the potential to improve decision making in precision agriculture by providing accurate predictive yield management methodology.

III. PROPOSED SYSTEM

3.1 EXISTING SYSTEM:

Leo Brieman's work focused on improving the accuracy and power of the random forest algorithm by exploring the interaction of its components. The algorithm works by creating multiple decision trees from different subsets of data and then combining their predictions through voting to produce the final output. Brieman proposed the use of a bootstrapping method to train the data and suggested that reducing the correlation among trees can improve the algorithm's accuracy without sacrificing its strength.

Balamurugan used only the random forest algorithm to predict crop yields by considering various factors such as rainfall, temperature, and time of year. However, the study did not compare the performance of other machine learning algorithms, making it difficult to determine the validity of the proposed method.

Mishra discussed various machine learning techniques in theory but did not apply any algorithms in practice, which limits the understanding of their functionality and effectiveness in real-world predictive environments.

DISADVANTAGES OF EXISTING SYSTEM:

1. The creation, repair, and maintenance of ML algorithms required huge costs as they are very complex.
2. The ML techniques used for crop yield prediction (such as KNN and linear regression) failed to provide accurate predictions in complex situations, such as extreme value data and nonlinear data.
3. Existing models often considered only one feature at a time for prediction, rather than considering all relevant features.
4. Inconsistent data from rainfall and temperature datasets can result in inconsistent predictions.

3.2 PROPOSED SYSTEM:

In this study, machine learning (ML) is utilized as a decision support tool for Crop Yield Prediction (CYP), aimed at aiding farmers in making informed decisions on crop selection and management during the growth season. Various ML approaches are combined to forecast crop output, and the results are evaluated based on mean absolute error. By incorporating factors such as temperature, rainfall, and area, among others, the predictions generated by ML algorithms can assist farmers in selecting the most appropriate crop to maximize yield.

Advantages of proposed system:

1. The proposed system combines multiple machine learning approaches, providing a more robust and accurate prediction of crop yield compared to relying on a single approach.

2. The use of variables like temperature, rainfall, and area in the prediction model allows for more comprehensive and informed decision-making for farmers.
3. The system can be used as a decision support tool for farmers, providing them with valuable insights to optimize crop selection and growth during the season.
4. The use of machine learning techniques provides more accurate predictions of crop yield, which can help farmers make better decisions on crop selection and management during the growing season.
5. By considering a variety of variables, such as temperature, rainfall, and area, the machine learning algorithms can provide more comprehensive and customized recommendations to farmers, resulting in increased crop yields and profits.
6. The use of machine learning techniques can reduce the reliance on traditional methods of predicting crop yields, which may not take into account all relevant variables and may be subject to human error.

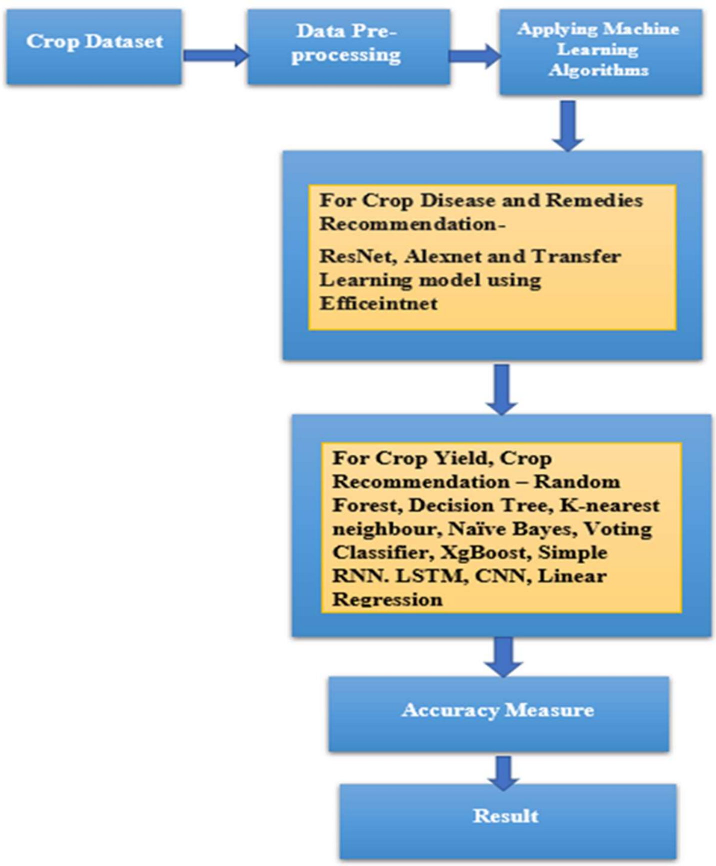


Fig1: Architecture of Proposed System

IV. STEPS FOR PROPOSED MODEL

4.1 MODULES:

Feature selection: Using this module, we will select the most relevant features for crop yield prediction and remedies recommendation, which will improve the accuracy of the model.

Data pre-processing: In this module, we will preprocess the data by cleaning, transforming, and normalizing the data to prepare it for modeling.

Model training: Using this module, we will train the selected machine learning algorithms on the preprocessed data to build a model for crop yield prediction and remedies recommendation.

Model evaluation: In this module, we will evaluate the performance of the trained models based on various evaluation metrics such as accuracy, precision, recall, and F1-score.

Remedies recommendation: Using this module, we will provide recommendations for remedies based on the predicted crop yield and other relevant factors such as soil type, weather, and irrigation.

Visualization: This module will be used to visualize the predicted crop yield and recommendations in a user-friendly format for better understanding and interpretation by the user.

4. 2 ALGORITHMS:

Each of the above algorithms can be used for crop yield prediction and remedies recommendation using feature selection techniques in machine learning. Here is a brief overview of how some of these algorithms can be applied:

1. Inception ResNet V2: Inception ResNet V2 can be used for crop yield prediction by training the model on a large dataset of crop images and associated yield information. The model can then be used to predict the yield of a new crop based on an image of the crop. Feature selection techniques can be used to identify the most important features in the images that are most strongly correlated with yield.
2. Random Forest: Random forest can be used for crop yield prediction by training the model on a dataset of crop features such as weather conditions, soil quality, and fertilizer use, along with associated yield information. The model can then be used to predict the yield of a new crop based on its features. Feature selection techniques can be used to identify the most important features that are most strongly correlated with yield.
3. Decision Tree: Decision tree can be used for crop yield prediction by training the model on a dataset of crop features and associated yield information. The model can then be used to predict the yield of a new crop based on its features. Feature selection techniques can be used to identify the most important features that are most strongly correlated with yield.
4. KNN: KNN can be used for crop yield prediction by training the model on a dataset of crop features and associated yield information. The model can then be used to predict the yield of a new crop based on its similarity to existing crops in the dataset. Feature selection techniques can be used to identify the most important features that are most strongly correlated with yield.
5. Naive Bayes: Naive Bayes can be used for crop yield prediction by training the model on a dataset of crop features and associated yield information. The model can then be used to predict the yield of a new crop based on its probability of belonging to a certain yield category.

Feature selection techniques can be used to identify the most important features that are most strongly correlated with yield.

6. Linear Regressor: Linear regressor can be used for crop yield prediction by training the model on a dataset of crop features and associated yield information. The model can then be used to predict the yield of a new crop based on its linear relationship with the features. Feature selection techniques can be used to identify the most important features that are most strongly correlated with yield.

7. XGBoost Regressor: XGBoost regressor can be used for crop yield prediction by training the model on a dataset of crop features and associated yield information. The model can then be used to predict the yield of a new crop based on its relationship with the features, using the XGBoost algorithm for gradient boosting. Feature selection techniques can be used to identify the most important features that are most strongly correlated with yield.

8. Voting Classifier: Voting classifier can be used for crop yield prediction by training multiple models on the same dataset of crop features and associated yield information. The models can then vote on the predicted yield of a new crop based on their individual predictions. Feature selection techniques can be used to identify the most important features that are most strongly correlated with yield.

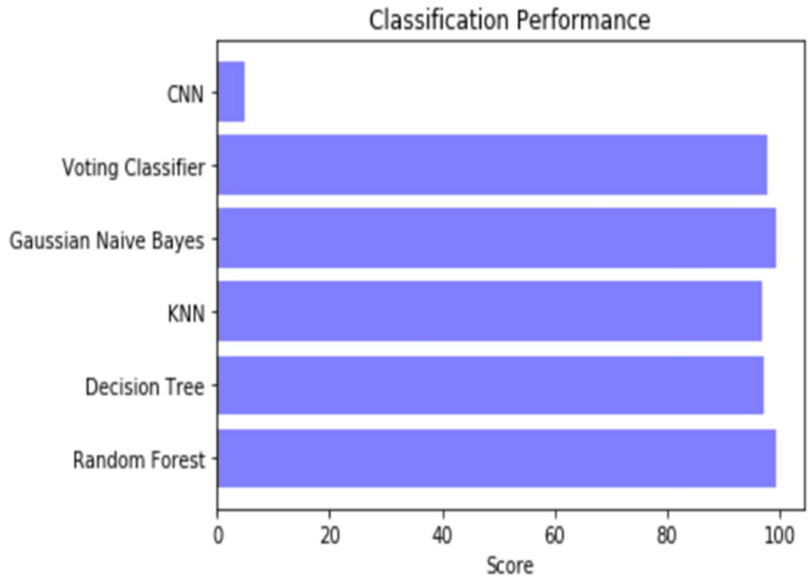
In summary, machine learning algorithms such as Inception ResNet V2, Random Forest, Decision Tree, KNN, Naive Bayes, Linear Regressor, XGBoost Regressor, and Voting Classifier can be applied for crop yield prediction and remedies recommendation using feature selection techniques. The choice of algorithm depends on the nature of the dataset and the specific problem at hand. Feature selection techniques can help identify the most important features that are most strongly correlated with yield, which can improve the accuracy of the predictions.

V. RESULT AND DISCUSSION

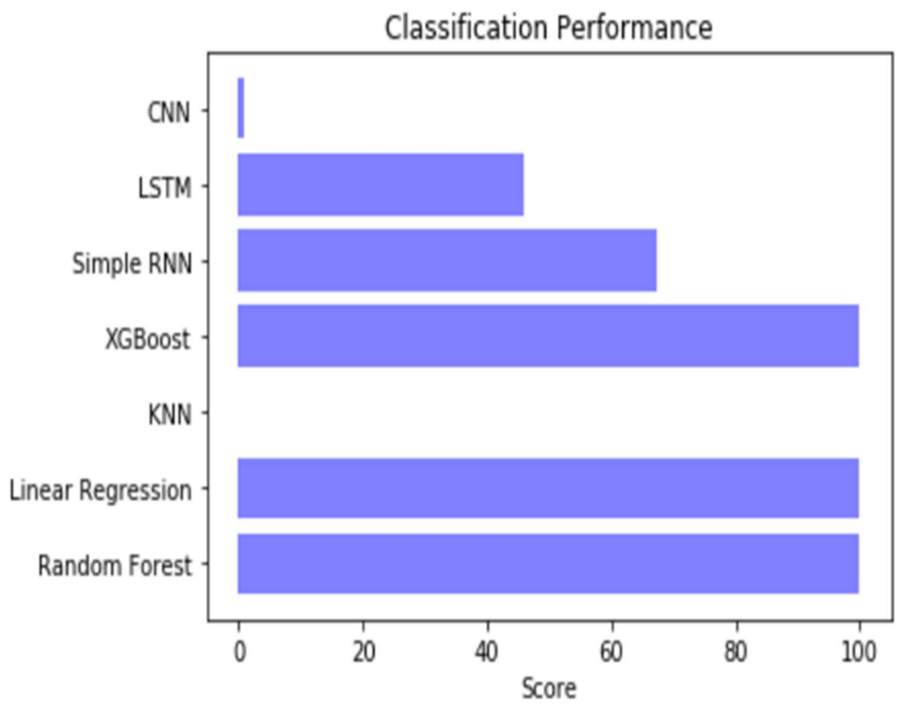
Machine Learning Algorithms	**Crop Recommendation Accuracy**	**Crop Yield Prediction Accuracy**	**Soil Moisture Accuracy**
Random Forest	99.3	99.8	98
Decision Tress	97.1	-	99.1
KNN	96.96	-	82
Gaussian Naive Bayes	99.44	-	75.5
Voting Classifier	97.7	-	98.3
Linear Regression	-	100	-
XgBoost	-	99.7	-
Simple RNN	-	67.35	-
LSTM	-	47.99	-
CNN	0.4	1.02	-

Result Graph

Recommendation Graph



Yield Prediction Graph



Soil Moisture Graph

VI. CONCLUSION

The study focuses on using machine learning (ML) algorithms as a decision support tool for Crop Yield Prediction (CYP) to assist farmers in choosing the best crops to grow for the greatest yield. The research explores various features, such as geological position, scale, and crop features, to determine the best performing features for CYP. Existing models using techniques such as neural networks, random forests, and KNN regression were investigated, but there is still room for improvement. The study proposes the use of ML in the agricultural domain field to advance crop prediction, and suggests further research into feature selection, such as considering temperature variation effects on agriculture. Additionally, fertilizer should also be considered in crop yield estimation, and a model based on deep learning (DL) should be developed for CYP. The study suggests focusing on treating delay to border topographical areas, using nonparametric models, and utilizing features from deterministic crop models to improve statistical CO2 fertilization. The study highlights the potential for ML to improve crop yield prediction and suggests avenues for future research to advance the field.

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] Mishra, S., Mishra, D., Santra, G. H.,Applications of machine learning techniques in agricultural crop production, Indian Journal of Science and Technology 10.17485/ijst/2016/v9i38/95032 volume 9,2016

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

[6] Ananthara, M. G., Arunkumar, T., & Hemavathy, R. (2013, February). CRY—an improved crop yield prediction model using bee hive clustering approach for agricultural data sets. In 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering (pp. 473- 478). IEEE.

[7] Awan, A. M., & Sap, M. N. M. (2006, April). An intelligent system based on kernel methods for crop yield prediction. In Pacific-Asia Conference on Knowledge Discovery and Data Mining (pp. 841-846). Springer, Berlin, Heidelberg.

[8] 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.

[9] Bhosale, S. V., Thombare, R. A., Dhemey, P. G., & Chaudhari, A. N. (2018, August). Crop Yield Prediction Using Data Analytics and Hybrid Approach. In 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA) (pp. 1-5). IEEE.

[10] Gandge, Y. (2017, December). A study on various data mining techniques for crop yield prediction. In 2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT) (pp. 420-423). IEEE.